REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Technical Report	

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Modeling Intercultural Collaboration and Negotiation	W911NF-08-1-0301
6. AUTHOR(S)	
See attached report	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Carnegie Mellon University 5000 Forbes Avenuek, Pittsburgh PA 15213	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING / MONITORING AGENCY REPORT NUMBER
U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	54235.2-LS-MUR

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT	12 b. DISTRIBUTION CODE
Approved for public release; distribution unlimited.	

13. ABSTRACT (Maximum 200 words)

14. SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU

W24:

MODELING INTERCULTURAL COLLABORATION AND NEGOTIATION (MICON)

Held in conjunction with IJCAI-09 (Twenty-first International Joint Conference on Artificial Intelligence), in Pasadena, CA on July 13, 2009

Workshop Organizers

Katia P. Sycara	Carnegie Mellon University
Michele Gelfand	University of Maryland
Allison Abbe	U.S. Army Research Institute

Preface

With increasing globalization of business and science, cultural differences of the parties are an important factor that affects the process and outcomes of collaborative and self-interested interactions.

The social science literature on culture as well as human collaboration and negotiation is vast. Most of this literature is devoted to work within the same culture. Artificial intelligence researchers, on the other hand, have developed computational models of cooperation, conflict resolution and negotiation, but paying almost no attention to identifying and modeling cultural factors. In recent years, we have witnessed a great increase in interest in understanding inter-cultural interactions. This has led to increased interest of social scientists and computational scientists in theoretical and experimental analysis of inter-cultural exchanges, modeling and support. Currently, these communities are largely unconnected. There is a great need to bring them together to share research work and experiences, discuss ideas and forge interdisciplinary collaborative relations.

This volume consists of proceedings of the first MICON (Modeling Intercultural Collaboration and Negotiation) workshop held in conjunction with IJCAI-09 (Twenty-first International Joint Conference on Artificial Intelligence), in Pasadena, CA on July 13, 2009. The workshop aims to bridge both computational and behavioral research on culture, collaboration and negotiation. We believe that the ongoing research presented in these proceedings will be of interest to researchers from AI/computer science/Economics and social/behavioral sciences fields, such as psychology, sociology, communications, organizational science.

The workshop organizers would like to thank all members of the Program Committee for their excellent work, effort, and support in ensuring the high-quality program and successful outcome of the first MCON workshop.

May 2009

Katia P. Sycara

Michele Gelfand

Allison Abbe

Organizing Committee

Katia P. Sycara	Carnegie Mellon University
Michele Gelfand	University of Maryland
Allison Abbe	U.S. Army Research Institute

Program Committee

Scott Atran	Anthropology, CNRS, France
Robert Axelrod	School of Public Policy, University of Michigan, USA
Geoff Gordon	Machine Learning, Carnegie Mellon University, USA
Toru Ishida	Social Informatics, University of Kyoto, Japan
Yoshi Kashima	School of Behavioral Science, University of Melbourne, Australia
Sarit Kraus	Math and CS, Bar Ilan University, Israel
Arie Kruglanski	Psychology, University of Maryland, College Park, USA
Michael Lewis	Information Sciences, University of Pittsburgh, USA
Dana Nau	Computer Science, University of Maryland, College Park, USA
Andrjez Nowak	Psychology, Florida Atlantic University and University of Warsaw, Poland
Avi Pfeffer	EECS, Harvard, USA
Eduardo Salas	Psychology, University of Central Florida, USA
Thomas Schelling	Economics, University of Maryland, College Park, USA
Catherine Tinsley	McDonough School of Business, Georgetown University, USA
Laurie Weingart	Tepper School of Business, Carnegie Mellon University, USA

Table of Contents

Towards a Theoretical Framework of Intercultural Collaboration

Wendy L. Bedwell, Jessica L. Wildman, Deborah DiazGranados,
Marissa Shuffler, Eduardo Salas, & C. Shawn Burke
Department of Psychology & Institute for Simulation and Training
University of Central Florida
3100 Technology Parkway
Orlando, FL 32826

Abstract. The purpose of this work is to address this gap in the literature by furthering the present understanding of multicultural collaborative units and the potential impact of cultural heterogeneity on collaboration. We propose a theoretical framework for examining multicultural collaboration grounded in multidisciplinary scientific theory. This framework addresses the situational, social and attitudinal constructs that impact the collaborative process. Moreover, we integrate an explanation of how individuals use their own cultural lens to interpret what occurs in their surroundings.

Keywords: Culture, Collaboration, etc.

1 Introduction

With increasing frequency, teams of teams are being called upon to make decisions and solve problems in complex, dynamic environments. Today, effective collaboration across departments, organizations, industries, and even globally across cultures, is critical to organizations and applied psychologists working to understand and develop teams. This has created a practical need to better understand collaboration in the context of multicultural interactions.

Yet to date, researchers have noted that there is no universally accepted definition of collaboration, nor unified, accepted model of what constitutes effective collaboration [1] Without a clearly defined term and a framework regarding the processes involved, readers are left to guess as to how collaboration is different from other constructs like cooperation or coordination, or even teamwork. Additionally this creates a problem for researchers in advancing a clear understanding of the state of the science on collaboration as it is not really clear what effective collaboration really is.

One major factor given today's global economy that has received far less attention than needed is the impact of culture on collaborative efforts. In civilian and military organizations alike, individuals from diverse cultural backgrounds are engaging in unparalleled collaborative efforts. Industrial globalization and the widespread use of

collaborative technologies to bridge time and distance gaps has created an environment that is conducive to utilization of teams of culturally diverse individuals distributed throughout the globe to address novel problems. Researchers have long noted the reciprocal nature of the relationship between culture and various psychological constructs. Yet, there has been little research examining the effects of cultural diversity on collaborative efforts.

As Lehman, Chiu, and Schaller [2] state, “Psychological processes influence culture. Culture influences psychological processes” (p. 689). Culture could have a significant impact on collective effectiveness as there are researchers who argue that individuals behave and think differently in accordance with their cultural influence [3]. Given these cross-cultural differences, the cultural composition of the collective could have a significant impact on team functioning and collaboration. Sutton, Pierce, Burke, and Salas [4] presented several propositions regarding the impact of cultural differences on team dynamics. These theorized differences in culture have the potential to affect all aspects of team functioning, including adaptive behavior.

Yet, to our knowledge, little, if any, research has looked at the specific impact of culture on effective collaboration. Questions such as those below still remain.

- What are the relevant cultural constructs?
- How does situational context affect collaboration across cultures?
- What factors facilitate success during a multicultural collaboration?
- How does culture affect social and collaborative processes?

In order to provide a foundation upon which to answer these fundamental questions, a theoretical framework is required. The purpose of this work is to address this gap in the literature by furthering the present understanding of multicultural collaborative units and the potential impact of cultural heterogeneity on collaboration. We propose a theoretical framework (see Figure 1) for examining multicultural collaboration grounded in multidisciplinary scientific theory. This framework addresses the situational, social and attitudinal constructs that impact the collaborative process. Moreover, we integrate an explanation of how individuals use their own cultural lens to interpret what occurs in their surroundings. Equally important in this framework is the role of leadership (or lack thereof) and its influence on individual members' cultural frames. In the following sections, we provide definitions of collaboration, culture, and a brief overview of our framework.

2 Defining Collaboration

We define collaboration as an evolving, macro process whereby two or more entities reciprocally engage in problem-solving activities to achieve mutually desired goals. This is based on underlying assumptions regarding collaboration which were derived from an integration of the extant multidisciplinary literature.

We assume that collaboration is not static. One of the most common reasons for parties to come together and collaborate is in response to some unpredictable, changing environment. Thus, collaboration is inherently dynamic in that all relationships change over time. Selin and Chavez [5] state that collaboration "is proposed to evolve dynamically in response to a host of internal and external factors" (p. 190).

Collaboration is a very high-level, "umbrella" term that encompasses many other types of interaction. One of the most salient themes found within the literature examining collaboration is that it cannot be confined to one particular level of analysis. Collaboration occurs between "individuals, groups, organizations, or even societies" [6].

Although some literature has conceptualized collaboration as a relationship structure, we agree with the majority of the multidisciplinary literature which has conceptualized collaboration as a process that parties engage in, rather than a structure [7,8]. Gray [7] describes notes that "collaboration is essentially an emergent process rather than a prescribed state of organization" (p. 15).

The most intuitive aspect of collaboration is that is involves working with others – notably, at least two or more entities – as evidenced by the fact that the dictionary definition of collaboration is "to work together"[9]. Consequently, for the sake of clarity and thoroughness we include the requirement of two or more entities as part of our definition.

Collaboration also requires both parties to be actively engaged in the collaborative process [6]. Essentially, one party dictating and controlling another party cannot be considered collaboration. The main advantage of collaboration versus more independent or individually-based problem-solving is the potential for sharing of resources, both tangible and intangible. We therefore assume that collaboration is a back-and-forth reciprocal process that requires each involved party to actively contribute in some meaningful way.

Collaboration is most often discussed in the context of intellectual endeavors and the creation of innovative and new knowledge. Therefore, we posit that collaboration is inherently focused on problem-solving. As we define problem-solving in the broadest sense, this does not necessarily limit collaboration to planning or decision-making tasks. It can also include more action-oriented tasks, which represent a more task-based form of problem-solving in that the involved parties must adapt to changing environmental factors, or "solve" environmental problems. For example, Graham and Barter [10] state that "above all else, collaboration captures the need to … develop innovative, new responses to rapidly changing … problems" (p. 6).

Finally, the existence of mutual, shared goals is likely the defining element separating collaboration from all other forms of shared work. Specifically, collaboration only exists if the involved parties share mutually agreed upon or mutually defined goals [10,11]. The shared goals are what make collaboration "collaborative." Without shared goals, there would be no reason for two or more entities to work together.

3 Defining Culture

We acknowledge that culture is a multi-faceted construct that has been defined by many. It has been defined as a shared set of individual ideas, behaviors, and products – which result from both ideas and behaviors [12]. Klein, Pongonis, and Klein [13] view culture consistent with an evolutionary perspective [14], that it has three defining characteristics: cultures are functional blueprints, dynamic systems, and have integrated components. Naylor [12] identifies three specific levels of cultural groupings: general culture, specific culture, and constituent culture. General culture refers to those characteristics that are uniquely human, specifically in the context of adaptation to the environment whereas specific culture is a more focused idea, behavior, or product related to a national or sociocultural context [12]. Continuing to narrow focus, Naylor defines a constituent culture as an idea, behavior, or product within one specific example of a sociocultural context, or a "subculture."

For the sake of clarity and the purposes of this paper we define culture as those organizational, team and individual level characteristics which are inherent or developed over time. These characteristics can take the form of values, ideas, norms, and behaviors. They are certainly dynamic, but most importantly, these characteristics are multi-level and are competing within an individual [12].

4 Intercultural Collaboration Framework

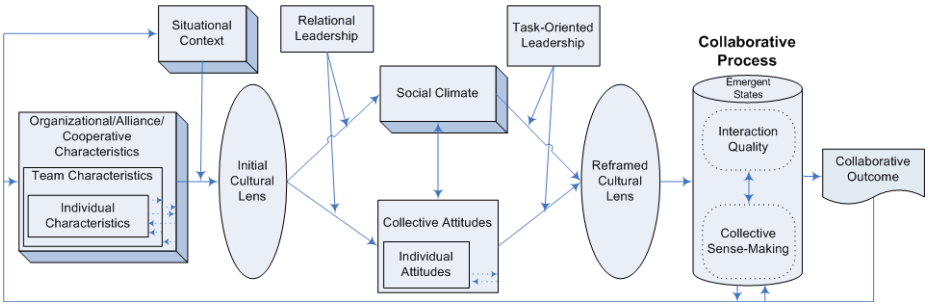

Fig. 1. Framework of Intercultural Collaboration

4.1 Inputs, Mediators, & Moderators

Our framework begins by capturing the dynamism of an individual, a team or an organization. Mathieu and colleagues [15] noted that individuals comprise teams, which work within organizations. The characteristics of these entities influence team effectiveness, yet also interact with each other (the characteristics of the individuals influence the characteristics of the team, etc.). An individual brings her own values to a team, the team can develop their own norms over time, and within an organization (or some higher level, this could be at the national or state level) there can exist a particular culture. Our framework incorporates these importance influences, noting their interdependent nature (see Figure 1). According to anthropology, each of these groups has their own existing culture [12]. Essentially, we conceptualize each of these inputs as existing cultures from which individuals draw upon to guide their behavior.

How do individuals who have never worked together resolve any competing characteristics/cultures? We propose that the situational context can moderate how they interpret their surroundings. The situational context can be made up of task characteristics (e.g., working with a close knit group who is not accepting of strangers) or the situation (e.g., extreme environment). It is then that the individuals of a collaborative unit use a cultural cognitive filter lens, in order to filter through which views and beliefs are translated into mechanisms with which to solve problems and make decisions. Klein [13] first conceptualized the notion of a cultural lens. We depict a cultural lens as being derived from the various individual, team, and collective characteristics/cultures, moderated by the situational context. It is the situation that determines which cultural characteristics will be most prescient to the individuals engaged in the collaborative effort. In other words, the culture which emerges in this collaborative unit can be any level of culture that has the most influence. Hence, it can be the culture found at the individual, team or organizational level – whichever culture has the most influence over behavior given the context.

This cultural lens will then influence the resulting social climate of the parties involved in the collaborative effort as well as the individual and collective attitudes. We define social climate consistent with the definition espoused by Pirola-Merlo and colleagues [16] as "the set of norms, attitudes, and expectations that individuals perceive to operate in a specific social context (p 564)." Individual attitudes are defined as an "internal state that influences an individual's choices or decisions to act in a certain way under particular circumstances" [17, p 352]; collective attitudes are defined based on the definition suggested by Salas, Goodwin, & Burke [18] as "internal states that are associated with the team and that affect the team's interaction processes (p 51)." We are not suggesting that the cultural lens *changes* these attitudes, it merely helps determine which of the existing individual and/or collective attitudes are most salient to the given collaborative effort.

It is at this time that we see the influence of the leader(s) during a collaborative effort. Before engaging in the collaborative process, individuals or teams who are interacting with one another must assess their social climate and collective attitudes. We propose that leadership, specifically person-oriented leadership or relational leadership has the most impact during the development of climate and attitudes. Judge and colleagues [19] defines

relational leadership as “the degree to which a leader shows concern and respect for followers, looks out for their welfare, and expresses appreciation and support” (p. 36). Furthermore, we believe that the practice of relational leadership helps to resolve conflict and problem solve in order to develop a more closely knit collaborative unit and a unit which has the attitude that they can accomplish the goal which they have been tasked with. As depicted in our framework, we also proposed that another form of leadership, task-oriented leadership, plays a large role in the collaborative unit just prior to the beginning of the collaborative process. For the sake of consistency, we again look to Judge et al. [19] for their definition of task-oriented leadership: “the degree to which a leader defines and organizes (her) role and the roles of followers, is oriented toward goal attainment, and establishes well-defined patterns and channels of communication” (p. 36). It is during this time that the leader may refocus the collaborative unit to the task at hand, and the processes which they are about to undertake. The resulting social climate and individual/collective attitudes in turn reframe the cultural lens. If a collaborative effort does not have a strong leadership (i.e. self-directed work teams) component, the cultural lens could remain largely unchanged.

4.2 Collaborative Process

The cultural lens exerts influence over the collaborative process. The collaborative process is made up both interaction quality and collective sense making. In order to describe the collaboration process we utilize Hoegl and Gemuenden’s [20] explanation of teamwork quality, and Weick’s [21] view of sense-making. Interaction quality consists of communication, coordination, balance of member contribution, mutual support, effort, and cohesive behaviors [20]. Sensemaking involves the ongoing retrospective development of plausible images that rationalize what people are doing. More simply put, sensemaking is effort to create order out of and interpret what occurs [21]. It is during this time that the collaborative unit is engaging in such processes such as communication, coordination, balancing member contribution, mutual support, effort, environmental scanning, and interpretation [20; 22; 23; 24; 25; 26; 27; 28).

Through a dynamic interaction between interaction quality and collective sensemaking, collaborative emergent states emerge such as shared mental models, shared identity, and mutual trust. This process allows the collaborative unit to organize members, assess their progress, and adapt as necessary in order to successfully accomplish their task and reach a collaborative outcome. To be clear, the result of the collaborative process can be a tangible or intangible product. This outcome does not only mean the products which are related to the overall goal for collaborating (e.g., balancing the national budget, restoring democracy to a country) but also those products that occur during the process and that can feed back into the framework of collaboration.

4.2 Collaborative Outcome

The collaborative process results in a collaborative outcome which can encompass a wide range of outcomes such as a tangible product, an idea, shared understanding, personal growth, or viability and satisfaction. While the collaborative process is occurring, emerging collaborative outcomes such as satisfaction and initial performance provide feedback to the collaborative process. As individuals and units within the collaboration receive feedback, changes in interaction quality and collective sense making may occur which in turn may improve or degrade the collaboration. The emerging and final collaborative outcomes can also alter the individuals, teams, and organizations within the collaboration, as well as the situational context. However, the feedback effects of the collaborative outcomes are expected to make less of an impact on the distal input factors compared to the impact it will have on the collaborative process. No clear end point is specified because collaborations can end upon the completion of the initial shared project or develop and continue indefinitely.

5 Conclusions and Implications

By developing this framework of intercultural collaboration and attempting to answer some fundamental questions, we hope that collaboration among members of different cultural subgroups will be better understood, which can lead to a better diagnosis of issues that impede successful collaboration. Additionally, we hope that this opens lines of multidiscipline communication to address important cultural issues such as the impact of culture on collaboration, as well as other team processes, in order to inform not only the science, but the practice.

References

1. Henneman, E. A., Lee, J. L., & Cohen, J. I. (1995). Collaboration: A concept analysis. *Journal of Advanced Nursing, 21(1),* 103-109.
2. Lehman, D. R., Chiu, C., & Schaller, M. (2004). Psychology and culture. *Annual Review of Psychology, 55,* 689-714.
3. Gibson, C. B., & Zellmer-Bruhn, M. E. (2002). Minding your metaphors: Applying the concept of teamwork metaphors to the management of teams in multicultural contexts. *Organizational Dynamics, 31,* 101-116.
4. Sutton, J. L., Pierce, L. G., Burke, C. S., & Salas, E. (2006). Cultural adaptability. In E. Salas (Series Ed.) & C. S. Burke, L. G. Pierce, & E. Salas (Vol Eds.) *Advances in*

human performance and cognitive engineering: Vol 6. A prerequisite for effective performance within complex environments, (pp. 143-173). Amsterdam: Elsevier.
5. Selin, S., & Chavez, D. (1995). Developing a collaborative model for environmental planning and management. *Environmental Management, 19(2)*, 189-195.
6. Longoria, R. A. (2005). Is inter-organizational collaboration always a good thing? *Journal of Sociology and Social Welfare, 32*(3), 123-138.
7. Gray, B. (1989). *Collaborating: Finding Common Ground for Multiparty Problems*. Jossey-Bass, San Francisco, CA.
8. Marttiin, P., Lehto, J. A., & Nyman, G. (2002). Understanding and evaluating collaborative work in multi-site software projects – A framework proposal and preliminary results. *Proceedings of the 35th Hawaii Conference on System Sciences.*
9. Keyton, J., Ford, D. J., & Smith, F. I. (2008). A mesolevel communicative model of collaboration. *Communication Theory, 18*, 376-406.
10. Graham, J. R., & Barter, K. (1999, January-February). Collaboration: A social work practice method. *Families in Society*, 6-13.
11. Gallant, M. H., Beaulieu, M. C., & Carnevale, F. A. (2002). Partnership: An analysis of the concept within the nurse-client relationship. *Journal of Advanced Nursing, 40(2)*, 149-157.
12. Naylor, L. L. (1996). *Culture and change: An introduction*. Westport, CT: Bergin & Garvey.
13. Klein, H.A., Pongonis, A., & Klein, G. (2000). *Cultural barriers to multinational C2 decision making*. Paper presented at the 2000 Command and Control Research and Technology Symposium.
14. Segall, M. H., Dasen, P.R., Berry, J.W., & Poortinga, Y.H. (1990). *Human behavior in global perspective: An introduction to crosscultural psychology.* New York: Pergamon.
15. Mathieu. J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team-effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management, 34*, 410-476.
16. Pirola-Merlo, A., Hartel, C., Mann, L., & Hirst, G. 2002. How leaders influence the impact of affective events on team climate and performance in R&D teams. *The Leadership Quarterly*, 13: 561-581.
17. Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo, E. Salas, & Associates (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco, CA: Jossey-Bass.
18. Salas, E., Goodwin, G. F., & Burke, C. S. (2009). *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches*. New York: Routledge.
19. Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology, 89*(1), 36-51.

20. Hoegl, M., & Gemuenden, H. G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science, 12*(4), 435-449.
21. Weick, K. E. (1995). Sensemaking in organizations. London: Sage.
22. Cartwright, D., & Zander, A. (1968). *Group dynamics: Research and theory* (2nd ed.). Evanston, IL: Row, Peterson.
23. Gioia, D. A., & Thomas, J. B. (1996). Identity, image, and issue interpretation: Sensemaking during strategic change in academia. *Administrative Science Quarterly, 41*(3), 370-403.
24. Hackman, R. H. (1987). The design of work teams. In J. L. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315-342). Englewood Cliffs, NJ: Prentice-Hall.
25. Pinto, M. B., & Pinto, J. K. (1990). Project team communication and cross-functional cooperation in new program development. *Journal of Product Innovation Management, 7*, 200-212.
26. Smircich, L., & Stubbart, C. (1985). Strategic management in an enacted world. *The Academy of Management Review, 10*(4), 724-736.
27. Tannenbaum, S. I., Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology, 43*, 399-441.
28. Thomas, J. B., Clark, S. M., & Gioia, D. A. 1993. Strategic sensemaking and organizational performance: Linkages among scanning, interpretation, action, and outcomes. *Academy of Management Journal***,** 36: 239-270.

A Formal Definition of Culture

Aliaksandr Birukou, Enrico Blanzieri, Paolo Giorgini, and Fausto Giunchiglia

DISI, University of Trento, via Sommarive 14, Povo(TN), Italy 38100
{birukou,blanzieri,pgiorgio,fausto}@disi.unitn.it,

Abstract. Globalization makes culture no more bound to a geographical area, race or religion. Multi-national companies, software developers, scientists need to take into account cultural differences when delivering products to people. The first step in dealing with culture consists in defining and representing culture of the targeted community. AI literature addressed issues of sociality, collaboration, and coordination in agent societies, but did not target the problem of defining and representing culture of a community. In this paper, we propose a formal definition of culture of a set of agents. It generalizes existing definitions of culture and it is operational in the sense that it can be applied for characterization and comparison of culture(s) existing in various communities.

1 Introduction

The advent of Web 2.0 lead to an explosive growth in the number of applications targeted at communities, e.g. applications supporting social navigation, collaborative editing, bookmarking and tagging. In such applications, culture is no more bound to a geographical area or a religion, as it is usually studied in anthropology. It becomes more appropriate to speak about the culture of online communities and such communities in general can not be characterized in terms of race, religion, or country. Rephrasing Axelrod [1], electronic communications allows us to develop patterns of interaction that are chosen rather than imposed by geography. Specific applications such as search engines or e-bookshops and the ways of using them become part of the culture of people. For instance, using Norton Commander file manager or preparing documents in the MS DOS 6.0 operating system, nowadays would be considered unusual to the same extent as lighting one's house with torches. Moreover, in some scenarios we can speak about societies of pure artificial agents, such as web services or programs and their specific culture, e.g. the standards implemented or the set of functionalities used. Human traders and trading agents operating on the same markets together use the same rules and develop common practices which can be referred to as culture. All this shows that grasping and representing culture becomes an important problem in computer science. Applications should be developed consistently with the culture of the target community and the notion of culture would provide support for building such applications.

AI literature on agents addresses the issues related to sociality, such as social action [2], social co-ordination architectures and social interaction strategies for

decentralized co-ordination in multi-agent systems [3], social laws and conventions in multi-agent environments [4, 5], and social roles [6]. However, the issue of sociality alone does not help neither to understand what differentiates one set of agents from another nor to grasp what are the specificities of the behavior of agents of a specific society. Although in two different agent societies agents can be able to communicate with each other and perform other social actions, these two societies can be very different from each other. We claim that the concept of culture can be used to describe and compare sets of agents. Some research approaches use the notion of culture in the context of agents, see e.g. [1, 7, 8], other provide a model for the comparison of cultures [9]. However, none of the previous research works provides a formal definition of culture that could be readily adopted for building applications for communities and applied for the characterization and the comparison of culture.

In this paper we provide a formal definition of culture. Our goal is not to provide a formalism or a reasoning framework per se, but, rather, to give an operational definition of culture that can be used for characterizing, describing, and comparing culture in different scenarios. In particular, we address the problems of development of applications according to the community culture and of characterizing culture of existing communities. We present and formalize a definition of the notion of culture of a set of agents. We define culture as a set of traits that are shared by the set and were transmitted, where traits are "characteristics of human societies that are potentially transmitted by non-genetic means" [10]. The sharing dimension is required for going from the set of personal traits of an individual to the culture of the set of agents, and to filter out such traits as divine services, marriage habits, birth rate, which only pertain to the set of agents as a whole, but not to individuals. The transmission dimension is a way of spreading culture.

The paper has the following structure: Section 2 reviews the use of concept of culture in the literature, Section 3 presents a formal definition of culture. Section 4 discusses related work and limitations of the approach, and Section 5 concludes the paper.

2 Defining culture

Culture is a slippery and ubiquitous concept. Initially, culture was associated with the notion of civilization tout-court. At the end of the 30s Margaret Mead put in contrast "culture" with "a culture". "Culture means the whole complex of traditional behavior which has been developed by the human race and is successively learned by each generation"([11] cited in [12]). However, specificity of the notion of culture with respect to a given human society was needed in order to study other societies. So the same citation goes on as: "A culture is less precise. It can mean the forms of traditional behavior which are characteristic of a given society, or of a group of societies, or of a certain race, or of certain area, or of a certain period of time" (cited in [12]). As a consequence, in the anthropological literature culture has been introduced as the concept denoting

the object of study of cultural anthropology. Other definitions were proposed and they largely vary. However, they seem to converge to the notion that culture is learned [1], it is associated with groups of people and its content includes a wide range of phenomena including norms, values, shared meanings, and patterned ways of behaving [13–18]. In anthropological literature the usefulness of the notion of culture as a scientific tool has been attacked giving rise to the so-called "writing against culture movement" (see Brumann [12] for a reaction against it). The culture as defined in anthropology usually refers to societies defined in national or ethnic terms, however, the concept of culture has been recently used for describing knowledge and behavior of other groups like in the concepts of corporate culture or organizational culture [13, 19, 20]. Moreover, globalization has brought about the problem of interaction of cultures. On the one hand, such interaction leads to blurring boundaries between cultures, while on the other hand it leads to the increasing need of cultural-aware managers and professionals. Recent anthropology textbook definitions take into account the shift in meaning as, for example, in the definition by Peoples and Bailey:

> *Culture is the socially transmitted knowledge and behavior shared by some group of people* (Peoples and Bailey [21, p. 23] cited in [12]).

Earlier authors define culture in the following ways (cited in Brumann [12]):

- Culture ... refers ... to learned, accumulated experience. A culture ... refers to those socially transmitted patterns for behavior characteristic of a particular social group (Keesing [22, p. 68]).
- Culture, or civilization, ... is that complex whole which includes knowledge, belief, art, law, morals, custom, and any other capabilities and habits acquired by man as a member of society (Tylor [23, p. 1]).
- The culture of any society consists of the sum total of ideas, conditioned emotional responses, and patterns of habitual behavior which the members of that society have acquired through instruction or imitation and which they share to a greater or less degree (Linton [24]).
- A culture is the total socially acquired life-way or life-style of a group of people. It consists of the patterned, repetitive ways of thinking, feeling, and acting that are characteristic of the members of a particular society or segment of a society (Harris [25]).

As we can see, definitions agree on the fact that culture consists of something that is shared and/or learned by a group of people, but the content of the culture varies in different definitions. Similarly to Axelrod [1], we see the content of the culture as a set of traits[1], which can refer to behavior, knowledge facts, ideas, beliefs, norms, etc.

[1] Traits are further grouped in *features* in Axelrod's formulation, i.e. each feature can take value from a set of specific traits.

3 A formal definition of culture

Consistently with AI literature, we define an *agent* as a "[...] physical or virtual entity that can act, perceive its environment (in a partial way) and communicate with others, is autonomous and has skills to achieve its goals and tendencies [...]" [26]. An agent can represent an individual or a collective entity such as an organization, and can have different *cultural traits*, which are characteristics of human societies that are potentially transmitted by non-genetic means and can be owned by an agent. The requirement "can be owned by", which we add to the definition by Mulder [10], means that it is possible for an agent to have a cultural trait. Different kinds of behavior, beliefs, knowledge, mentioned as elements of culture previously, are just particular kinds of cultural traits in terms of our formalism.

To model changes in the set of traits of an agent and consequently, changes in culture, we use the notion of state. We assume that the world can be in different states and the set of traits of the same agent can be different in different states.

Let us consider the set of agents Ag, the set of traits $\mathcal{T}$, and the set of states S. Given an agent $a \in Ag$ and a state $s \in S$, we denote the set of cultural traits of the agent a in the state s with $T_a(s) = \{\tau_i\} \subseteq \mathcal{T}$ and we use the predicate $has(a, \tau_i, s)$ to represent the fact that the agent a has a trait $\tau_i \in T_a(s)$ in the state s. In the following, we call the set of traits of an individual *the culture of an individual*.

Example 1. Let us consider a set of people and model them as agents with a set of traits and some behavior related to transmission, in particular, *telling* and *memorizing*. Let Ag in our example be a set of people: Charlie, Pedro, Maria, and Andrea are European citizens, and Toru is from Japan. Let $\mathcal{T}$ be a set of traits of different types, as shown in Table 1. For each trait, we also put its abbreviation (used in the figures later) in parentheses.

trait type	**traits**
knowledge	$Dante\_Alighieri\_wrote\_Divine\_Comedy(DA)$, $cappuccino\_is\_coffee(CI)$, $latte\_macchiato\_is\_coffee(LM)$, $Meiji\_era\_was\_in\_1868\_1912(ME)$
behavior	$eating\_with\_sticks(ES)$, $telling$, $memorizing$, $eating\_with\_fork(EF)$
norms, rules	$never\_put\_mayonnaise\_on\_pizza(NP)$, $never\_open\_umbrella\_inside\_building(NO)$
beliefs	$Christianity(Chr)$, $Buddhism(Bud)$

Table 1. The set of traits $\mathcal{T}$ in Example 1.

Table 2 lists the sets of traits of the specific agents of $Ag = \{Charlie, Pedro, Toru, Maria, Andrea\}$ in the state s_1. We can write $has(Maria, Dante\_Alighieri\_wrote\_Divine\_Comedy, s_1)$, or $has(Charlie, cappuccino\_is\_coffee, s_1)$, but not $has(Andrea, eating\_with\_sticks, s_1)$. We will use this example as a running example. ⊘

set	traits
$T_{Charlie}(s_1)$	$Dante\_Alighieri\_wrote\_Divine\_Comedy$, $latte\_macchiato\_is\_coffee$, $telling$, $cappuccino\_is\_coffee$, $eating\_with\_sticks$, $eating\_with\_fork$, $never\_put\_mayonnaise\_on\_pizza$, $Buddhism$
$T_{Pedro}(s_1)$	$Dante\_Alighieri\_wrote\_Divine\_Comedy$, $latte\_macchiato\_is\_coffee$, $cappuccino\_is\_coffee$, $eating\_with\_fork$, $Christianity$
$T_{Toru}(s_1)$	$Meiji\_era\_was\_in\_1868\_1912$, $cappuccino\_is\_coffee$, $eating\_with\_sticks$, $Buddhism$, $memorizing$
$T_{Maria}(s_1)$	$Dante\_Alighieri\_wrote\_Divine\_Comedy$, $latte\_macchiato\_is\_coffee$, $cappuccino\_is\_coffee$, $eating\_with\_sticks$, $eating\_with\_fork$, $Christianity$
$T_{Andrea}(s_1)$	$Dante\_Alighieri\_wrote\_Divine\_Comedy$, $latte\_macchiato\_is\_coffee$, $cappuccino\_is\_coffee$, $eating\_with\_fork$, $Christianity$

Table 2. Traits of agents in Example 1.

Note that we do not introduce types of traits and use them in the example only for convenience. One might propose a different classification of traits, e.g. putting *eating_with_sticks* as a norm. We believe that there is no single classification and it is better to deal with generic traits rather than with specific types of cultural content.

We distinguish behavior as a particular kind of traits and assume that performing a behavior by an agent changes the state of the world. In line with AI literature, we define *behaviors* as "[...] reified pieces of activity in which an agent engages, for example sleep or eat. In colloquial English an agent behaves in various ways; in technical AIese, an agent has various behaviors" [27]. We define the set of all behaviors $\mathcal{B} \subseteq \mathcal{T}$ and the function $perform$ in $Ag \times \mathcal{B} \times S \to S$. The intended meaning of this function is that an agent, which has some behavior in some state, performs this behavior in this state and the state of the world changes to another state. More specifically, $s_v = perform(a, \tau, s_u)$ means that $has(a, \tau, s_u)$ and the agent a performed a behavior τ in the state s_u and the resulting state is s_v. The fact that $has(a, \tau, s_u)$ does not imply that the agent a is able to perform the behavior τ in the state s_u, because some preconditions for performing the behavior may be not fulfilled in the state s_u. Note that since traits are not innate, by assuming $\mathcal{B} \subseteq \mathcal{T}$ we do not include innate behaviors, such as blinking when air is puffed in someone's eye.

At this point we would like to discuss the distinction between *action* and *behavior*. In AI literature, an action is an atomic piece of activity, while behavior is perceived as something more complex, and can include several actions. Therefore, our notion of performing a behavior can really be decomposed into performing several actions. However, we decided not to introduce explicit relations between actions and behaviors. Moreover, the absence of such clear dependency in AI literature suggests that these relations are hard or even impossible to formalize. Instead, we assume that behavior can represent an atomic action or a more complex activity depending on the level of modeling granularity. We can

vary granularity of behaviors depending on the problem in hand and on the domain. For instance, in Example 1, when someone needs to know whether agents are working, it is possible to consider behaviors *working* and *eating*, or, even, *working* and *not_working*. However, if someone would like to have a closer look at eating habits of the group, it is necessary to introduce finer granularity of the *eating* behavior, e.g. by considering *eating_with_sticks* and *eating_with_fork* behaviors.

We assume that the states are ordered, we define recursively the order "is before" and the corresponding predicate $is\_before(s_u, s_v)$ and $is\_after(s_v, s_u)$ in the following way:

Definition 1 (is_before). $is\_before(s_u, s_v) \leftrightarrow \exists a \in Ag,\ \tau \in \mathcal{B},\ s \in S$ *such that* $s = perform(a, \tau, s_u) \wedge (s = s_v \vee is\_before(s, s_v))$.

Definition 2 (is_after). $is\_after(s_v, s_u) \leftrightarrow is\_before(s_u, s_v)$

We assume that in each state s_v, the previous state s_u is uniquely defined, while the next state depends on the action an agent performs in s_v. We also state the following axiom:

Axiom 1 *For all agents* $a \in Ag$*, for all behaviors* $\tau \in \mathcal{B}$ *and for all states* $s_u, s_v \in S$

$$s_v = perform(a, \tau, s_u) \rightarrow is\_before(s_u, s_v)$$

Definition 3 (sharing). *For each pair of agents* $a_i, a_j \in Ag$*, for each trait* $\tau \in \mathcal{T}$*, and for each state* $s \in S$*,* a_i *and* a_j *share the trait* τ *in the state* s *iff they both have such a trait in* s*:*

$$has(a_i, \tau, s) \wedge has(a_j, \tau, s) \leftrightarrow sharing(a_i, a_j, \tau, s).$$

We also assume that agents do not lose traits when the state of the world changes, as the following axiom says:

Axiom 2 *For all agents* $a \in Ag$*, traits* $\tau \in \mathcal{T}$*, and states* $s \in S$ *:*

$$has(a, \tau, s) \rightarrow \forall s_v : is\_after(s_v, s)\ has(a, \tau, s_v).$$

Example 1 (continued). In the example, we can write $sharing(Toru, Maria, eating\_with\_sticks, s_1)$, or $sharing(Pedro, Andrea, cappuccino\_is\_coffee, s_1)$, etc. To avoid giving the complete list of tuples for which *sharing* holds, we represent them as a graph where nodes are agents and labels on each edge denote traits that are shared by the pair of agents connected by the edge, see Figure 1 for the state s_1. ⊘

Let us assume that if an agent a_i has a trait τ, the trait τ can be transmitted to another agent a_j before some state s and we use the predicate $transmitted(a_i, a_j, \tau, s)$ to represent this. We represent $transmitted(a_i, a_j, \tau, s)$ in a graph by a directed edge from a_i to a_j labeled τ.

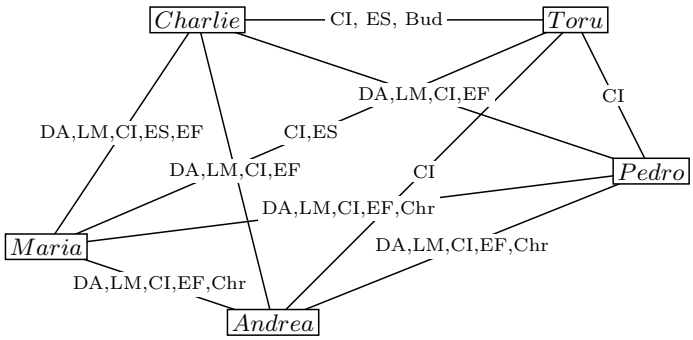

Fig. 1. The graph showing for which agents and traits the predicate *sharing* holds in Example 1. The nodes are agents and labels on each edge denote traits that are shared by the pair of agents connected by the edge. For instance, the edge between $Toru$ and $Andrea$ labeled CI means that $sharing(Andrea, Toru, cappuccino\_is\_coffee)$. The traits are abbreviated as in Table 1: $Dante\_Alighieri\_wrote\_Divine\_Comedy$ is abbreviated as DA, $latte\_macchiato\_is\_coffee$ as LM, $cappuccino\_is\_coffee$ as CI, $eating\_with\_sticks$ as ES, $eating\_with\_fork$ as EF, $Christianity$ as Chr, $Buddhism$ as Bud.

Definition 4 (transmitted). *For each pair of agents $a_i, a_j \in Ag$, $a_i \neq a_j$, for each trait $\tau \in \mathcal{T}$, and for each state $s \in S$ we say that the trait τ has been transmitted from a_i to a_j before the state s iff exists some state $s_u \in S$ such that a_i has τ in the state s_u, a_j does not have τ in the state s_u and an agent a_k performing a behavior τ_m in the state s_u imply that in the resulting state s_v the agent a_j has τ:*

$$(\exists s_u \in S,\ is\_before(s_u, s)\ has(a_i, \tau, s_u) \wedge \neg has(a_j, \tau, s_u) \wedge (s_v = perform(a_k, \tau, s_u)) \rightarrow has(a_j, \tau, s_v)) \leftrightarrow transmitted(a_i, a_j, \tau, s)$$

From our assumption that traits are not innate, it follows that traits are acquired by agents, and the goal of the transmitted predicate is to show the way an agent acquired a trait. For the sake of the expressivity of the model, we assume that in the initial state agents have some traits and the way they acquire other traits is represented using the transmitted predicate.

We should note that the trait τ is not shared by a_i and a_j in the state s_u, while it is shared by a_i and a_j in the state s_v, and in the state s, as shown by the following property:

Property 1. For all pairs of agents $a_i, a_j \in Ag$, for all traits $\tau \in \mathcal{T}$, and for all states $s_v \in S$

$$sharing(a_i, a_j, \tau, s_v) \rightarrow (\forall s : is\_after(s, s_v)\ sharing(a_i, a_j, \tau, s))$$

From Definition 4 it also follows that the transmitted predicate holds for all subsequent states after s_v.

Property 2. For all pairs of agents $a_i, a_j \in Ag$, for all traits $\tau \in \mathcal{T}$, and for all states $s_v \in S$

$$transmitted(a_i, a_j, \tau, s_v) \rightarrow (\forall s : is\_after(s, s_v)\ transmitted(a_i, a_j, \tau, s))$$

Example 1 (continued). Figure 2 shows the graph representing the *transmitted* predicate in state s_1 in our example. The traits *Dante_Alighieri_wrote_Divine_Comedy* and *eating_with_sticks* have been transmitted. On the contrary, the traits *cappuccino_is_coffee* and *never_put_mayonnaise_on_pizza* have not been transmitted (the latter trait is not even shared by any pair of agents). In particular, the *Dante_Alighieri_wrote_Divine_Comedy* trait has been transmitted from *Charlie* to *Maria*, and from *Maria* to *Andrea*. Also, the *eating_with_sticks* trait has been transmitted from *Charlie* to *Toru* and from *Toru* to *Maria*. We can write *transmitted*(*Charlie*, *Maria*, *Dante_Alighieri_wrote_Divine_Comedy*, s_1).

Let us assume that in the state s_1 *Charlie* tells *Toru* that Dante Alighieri wrote the Divine Comedy. In the next state, s_2, *Toru* memorizes this piece of knowledge. This corresponds to $s_2 = perform(Charlie, telling, s_1)$ and $s_3 = perform(Toru, memorizing, s_2)$. The *transmitted* predicate in the state s_2 is as depicted in the left part of Figure 2 and *transmitted* in the state s_3 is as depicted in the right part of Figure 2. The difference in the *transmitted* predicates in these two states is that the *Dante_Alighieri_wrote_Divine_Comedy* trait has been transmitted from *Charlie* to *Toru* and the corresponding edge is added, namely *transmitted*(*Charlie*, *Toru*, *Dante_Alighieri_wrote_Divine_Comedy*, s_3). Let us also assume that in the state s_2 the set of traits for each agent is the same as in the state s_1, while in the state s_3 the following change occurs: $T_{Toru}(s_3) = \{Meiji\_era\_was\_in\_1868\_1912,\ Dante\_Alighieri\_wrote\_Divine\_Comedy,\ cappuccino\_is\_coffee,\ eating\_with\_sticks,\ Buddhism,\ memorizing\}$.

Obviously, the transmission has an impact on sharing and the *sharing* predicate in the state s_3 is as depicted in Figure 3, with the edges between *Toru* and *Charlie*, *Maria*, *Andrea*, *Pedro* added. ⊘

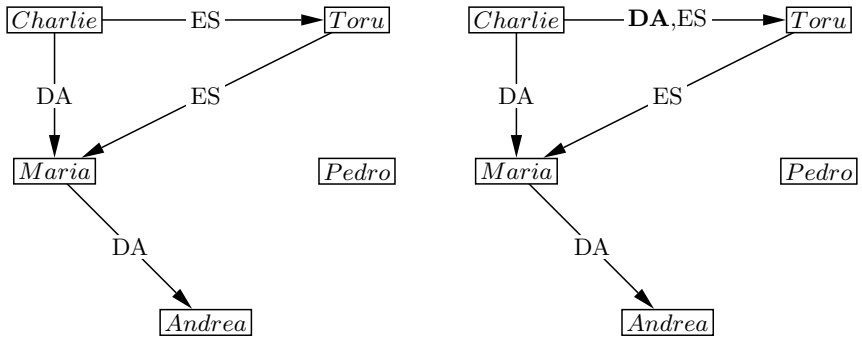

Fig. 2. The graph that shows for which agents the *transmitted* predicate holds in the state s_1 (left) s_3 (right) in Example 1. Changes with respect to state s_1 are in bold.

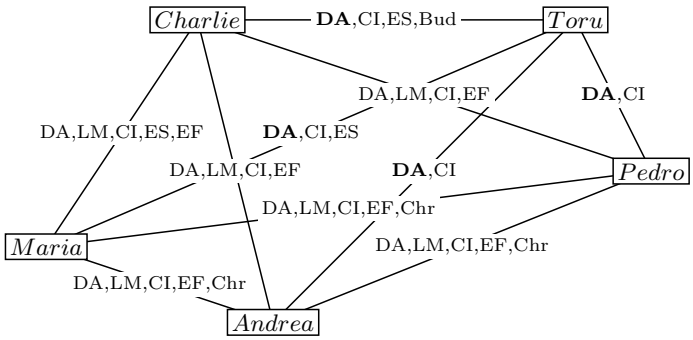

Fig. 3. The graph that shows for which agents the *sharing* predicate holds in the state s_3 in Example 1. Changes with respect to state s_1 are in bold.

Given a set of agents $G \subseteq Ag$ and a set of traits $T_G \subseteq \mathcal{T}$ we define the notions of *weak sharing* and *strong sharing*.

Definition 5 (weak sharing). *A set of traits T_G is weakly shared by a set of agents G in a state s iff for each trait $\tau \in T_G$ there exists a pair of agents $a_i, a_j \in G$, $a_i \neq a_j$ that share τ in the state s.*

Definition 6 (strong sharing). *A set of traits T_G is strongly shared by a set of agents G in a state s iff each trait $\tau \in T_G$ is shared by all pairs of agents $a_i, a_j \in G$ in s.*

In other words, the set of traits is weakly (strongly) shared if it is a subset of the union (intersection) of traits shared by pairs of agents of G in the state s.

Example 1 (continued). Let us consider the set of agents $G = \{Charlie, Toru, Maria, Andrea, Pedro\}$. Analyzing the *sharing* predicate in the state s_1 (Figure 1) we can see that only the $cappuccino\_is\_coffee$ trait is shared by each pair of agents in the state s_1, so $T_G = \{cappuccino\_is\_coffee\}$ is strongly shared by G in the state s_1. There are three traits that are shared by at least one pair of agents in the state s_1: $cappuccino\_is\_coffee$, $eating\_with\_sticks$ shared, e.g, by $Toru$ and $Charlie$, and $Dante\_Alighieri\_wrote\_Divine\_Comedy$ shared, e.g., by $Charlie$ and $Andrea$. So, the set $T'_G = \{Dante\_Alighieri\_wrote\_Divine\_Comedy, cappuccino\_is\_coffee, eating\_with\_sticks\}$ and all non-empty subsets of this set are weakly shared by the set G in the state s_1. Analogously, the set $T''_G = \{eating\_with\_sticks, Dante\_Alighieri\_wrote\_Divine\_Comedy, cappuccino\_is\_coffee\}$ is weakly shared by G in the state s_3, and the set $T'''_G = \{cappuccino\_is\_coffee, Dante\_Alighieri\_wrote\_Divine\_Comedy\}$ is strongly shared by the set G in the state s_3. ⊘

Property 3. Strong sharing implies weak sharing.

Given a set of agents $G \subseteq Ag$ such that $|G| \geq 2$, and a *transmitted* predicate we introduce the notion of culture of G.

Definition 7 (weak culture of a set of agents). *A non-empty set of traits $T_G \subseteq \mathcal{T}$ is a weak culture of G in a state s iff*

- *the set T_G is weakly shared by G in the state s,*
- *for each agent $a \in G$ in the state s there exists a trait $\tau \in T_G$ such that $has(a, \tau, s)$.*

From the assumption that traits are not innate, as we discussed, it follows that traits are acquired by agents, as represented by the transmitted predicate. Therefore, we can formulate the following axiom, telling that all traits in culture are transmitted.

Axiom 3 *For each trait $\tau \in T_G$ there exists an agent $a \in Ag$ that transmitted τ to another agent $a_j \in G$before the state s, i.e. $transmitted(a, a_j, \tau, s)$.*

From Definition 7 and Axiom 3 it follows that all the traits in the culture are transmitted, shared, and each agent has at least one trait from the culture. Please, note that since the traits are transmitted not necessarily within the set, the transmitted predicate does not imply sharing between the agents of G.

Definition 8 (strong culture of a set of agents). *If T_G in Definition 7 is also strongly shared in the state s then it is* a strong culture of the set of agents *G in the state s.*

In the following if we refer to "a culture of a set of agents", we mean "a weak culture of a set of agents".

Example 1 (continued). Considering $G = \{Toru, Andrea\}$ in the state s_3, $T_G = \{Dante\_Alighieri\_wrote\_Divine\_Comedy,\ cappuccino\_is\_coffee\}$ is strongly shared by the set G in the state s_3.

Although the $Dante\_Alighieri\_wrote\_Divine\_Comedy$ trait has been transmitted both to $Toru$ and $Andrea$ from outside (from $Charlie$ and $Maria$, respectively), it is strongly shared by the agents of G. Since in the state s_3 each agent in G has the trait $Dante\_Alighieri\_wrote\_Divine\_Comedy$, $T'_G = \{Dante\_Alighieri\_wrote\_Divine\_Comedy\}$ is a culture of G in the state s_3. It is easy to see that T'_G is not a culture of G in the states s_1 and s_2 because $Toru$ does not have this trait in those states. ⊘

The following proposition outlines some restrictions on how culture can change between states, namely it shows that culture is monotonic.

Proposition 1 (monotonicity of culture). *If a non-empty set of traits T_G is a culture of a set of agents G in a state s_v, then T_G is a culture of G also in any state s after s_v.*

In real world, the traits of a culture can be lost for two reasons: (1) agents can lose traits, (2) agents can die, move to another group, etc. As we stated in Axiom 2, in our model, agents do not lose traits. However, our model, and the proposition about monotonicity of culture support the case when agents disappear from the group.

Definition 9 (union culture of a group). *A non-empty set of traits T_G^{union} is the union culture of a set of agents G in the state s iff T_G^{union} is the union of all cultures T_G of G in the state s.*

In other words, the union culture of a set of agents in some state is the union of all possible cultures of the set in this state. Since it is the union of all cultures, it is not possible to add any trait to T_G^{union} and still obtain a culture of G. In the following, we refer to the union culture of a set of agents as "the culture of a set".

Definition 10 (evolution of culture). *A sequence of sets of traits $\{T_G^{(1)}, \ldots, T_G^{(i)}\}$ is an evolution of culture of G iff:*

- *exists a sequence of states $\{s_1, \ldots, s_i\}$, such that $T_G^{(k)}$ is a culture of G in the state s_k for all k, $1 \leq k \leq i$,*
- *for each k, $1 \leq k \leq i-1$ holds $is\_after(s_{k+1}, s_k)$.*

In other words, a sequence of sets of traits is an evolution of culture if each set of traits in the sequence is a culture of G in some state and the states are ordered in the same way as the sets of traits. We denote evolution of culture as $\{T_G\}$.

4 Related work and discussion

Carley [7] considers culture as the distribution of information (ideas, beliefs, concepts, symbols, technical knowledge, etc.) across the population and proposes a model for knowledge transfer based on interactions. In that model, the probability of an interaction between two agents is based on the principle of homophily, i.e. the greater the amount of knowledge they share the more probable the interaction is. During an interaction, agents exchange facts, so after the interaction one of the agents might know more than before the interaction. The knowledge transfer in these settings can be seen as a particular kind of culture spread. This work is further extended in the Construct project [28, 29]. For instance, one of the recent applications of Construct studies the effects of different methods of information diffusion on spreading beliefs and knowledge about illegal tax schemes in different American cities [30]. With respect to the definition of culture we propose in this paper, that model of information diffusion is complementary, because it models transmission of elements of culture (e.g., beliefs, knowledge) in a society.

Axelrod [1] considers culture as a list of features or dimensions. Each feature represents an individual attribute that is subject to social influence and can have different values called traits. Two individuals have the same culture if they have the same traits for all features. Similarly to the work by Carley, feature of an agent can change its value during an interaction and the probability of interaction is based on the homophily.

The notion of trait we use in our formalism is similar to the notion of feature used by Axelrod, specifically, each feature can take value from a set of specific

traits. Traits in our formalism also includes ideas, beliefs and technical knowledge used as culture elements by Carley. Both theories by Carley and by Axelrod are based on the assumption that culture changes as a result of an interaction. Thus, in our terms, interaction in that sense can be considered as a particular kind of transmission: there are two agents participating, it takes place in some specific state and it leads to the appearance of some cultural element in one of the agents.

Epstein and Axtell [8] study the emergence of the group rules from local ones defined at an agent's level in an artificial society of simple agents living and consuming sugar in an artificial environment called "Sugarscape". The authors consider a culture of the society as a string of binary cultural attributes and model cultural transmission both on horizontal (between agents) and vertical (through generations) levels using simple rules. However, they do not provide any formal definition of culture since the main focus of the book is on the emergence of group rules from the local ones.

According to O'Reilly [13], the culture of an organization is considered as strong if wide consensus exists about the content and participants believe in the importance of the content. They also formulate this as a [not necessarily big] set of values that are widely shared and strongly held. This is similar to the notion of strong culture, i.e. culture shared by all pairs of agents in a group, we consider in our formalism.

Balzer and Tuomela [31] study social practices and the dynamics of their maintenance in groups. They define social practices as recurrent collective activities based on collective intentions. The paper focuses on informal, non-normative practices, such as playing soccer on Sundays, going to sauna on Saturday afternoon, shaking hands, sharing a ride to work. They also note that the maintenance (change, preservation, renewal) depend on the success of a practice. The main contribution of the paper is a mathematical model for the description of social practices and their maintenance in groups.

Our model of culture is not limited to social practices. Moreover, it allows for inclusion of normative practices as well. However, as a consequence, the model of Balzer and Tuomela allows for a richer description of informal social practices. For instance, our model does not permit expressing intentions, but allows operating on manifestations of activities without going into details of underlying intentions. While authors show that success of a social practice is important for its adoption, for our model it is irrelevant whether a trait is successful in some sense. Our model just captures the fact that the trait is a part of culture, no matter how it occurred. The model presented by Balzer and Tuomela is defined for groups and then goes to the individual level, thereby implementing top-down approach. In our model of culture, we start from a set of traits of an individual, consider transmission as an important means of spreading culture, and then go to the culture of a group. Thus, we implement bottom-up approach. Balzer and Tuomela, while requiring sharing of a social practice within a group, and noting the importance of transmission for spreading practice, include transmission into the model only to a certain extent, namely, considering imitation as an example

of transmission. Our model of culture allows for different types of transmission as long as there is a predicate that helps to distinguish occurred transmissions.

Hofstede [9] treats culture as "[...] the collective programming of the mind that distinguishes the members of one group or category of people from another", proposes a model of culture and applies it for studying and comparing cultures of IBM workers in more than 50 countries. The model includes the following five independent dimensions of national culture differences: *power distance*, which is related to the different solutions to the basic problem of human inequality; *unvertainty avoidance*, which is related to the level of stress in a society in the face of an unknown future; *individualism* versus *collectivism*, which is related to the integration of individuals into primary groups; *masculinity* versus *femininity*, which is related to the division of emotional roles between men and women; and *long-term* versus *short-term* orientation, which is related to the choice of focus for people's efforts: the future or the present. Values in Hofstede's terms refer to "a broad tendency to prefer certain states over others" and are similar to attitudes and beliefs, which are just particular kind of traits in our formalism. Dimensions, similarly to Axelrod's features, take values from the set of traits. Thus, comparing with our work, the model developed by Hofstede has a different focus - it aims at comparing cultures of groups of people over several pre-defined dimensions of values, while our model supports comparison over arbitrary sets of traits. The dimensions in Hofstede's model are meant to be independent, while our formalism does not address the issue of dependency of traits, so they can be dependent on each other. In this line of thoughts, an interesting application of our model could be comparison of dependency of traits across groups, i.e. if presence of a trait or traits leads to the presence of another trait(s) for one group and to the presence of third trait(s) for another group.

The definition of culture presented here allows for representation and comparison of different cultures. However, in order to compare traits, one first needs to identify the traits of individuals. On the one hand, deducing traits from manifested behaviors of agents is not a trivial task in general. On the other hand, in specific domains this might be much easier, consider, for instance, deducing traits of users from logs of a web service, website, or an application. For instance, it would be possible to see that a group of users of a text editor always turn off the autocorrect feature and turn it off automatically in new versions of the editor prepared for this group. Taking the issue of the observability of traits into account, we see social software and Web 2.0 systems as one of the potential application domains for our model.

5 Conclusion

In this paper we have presented a formal definition of culture of a set of agents. This definition addresses existing gaps in AI literature that deals with issues of sociality, cooperation, and negotiation, but remains oblivious to the notion of culture. The formalism presented in this paper is a part of ongoing research and

we are currently working on measures for characterizing community culture and on studying of evolution of culture in Web 2.0 communities.

Acknowledgments

This work has been supported by the European Union FP7 projects COMPAS and LiquidPublication, grants no. 215175 and 213360.

References

1. Axelrod, R.: The dissemination of culture: A model with local convergence and global polarization. The Journal of Conflict Resolution **41**(2) (1997) 203–226
2. Castelfranchi, C.: Modelling social action for ai agents. Artificial Intelligence **103**(1-2) (1998) 157–182
3. Ossowski, S.: Co-ordination in Artificial Agent Societies, Social Structure and Its Implications for Autonomous Problem-Solving Agents. Volume 1535 of Lecture Notes in Computer Science. Springer (1999)
4. Shoham, Y., Tennenholtz, M.: On social laws for artificial agent societies: off-line design. Artif. Intell. **73**(1-2) (1995) 231–252
5. Shoham, Y., Tennenholtz, M.: On the emergence of social conventions: modeling, analysis, and simulations. Artificial Intelligence **94**(1-2) (1997) 139–166
6. Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., Guarino, N.: Social roles and their descriptions. In: Principles of Knowledge Representation and Reasoning (KR2004). (2004)
7. Carley, K.: A theory of group stability. American Sociological Review **56**(3) (1991)
8. Epstein, J.M., Axtell, R.L.: Growing Artificial Societies: Social Science from the Bottom Up. The MIT Press (1996)
9. Hofstede, D.G.: Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. 2nd edn. Sage Publications, Inc (April 2001)
10. Mulder, M.B., Nunn, C.L., Towner, M.C.: Cultural macroevolution and the transmission of traits. Evolutionary Anthropology **15**(2) (2006) 52–64
11. Mead: Cooperation and Competition Among Primitive Peoples. Transaction Publishers (2002)
12. Brumann, C.: Writing for culture: Why a successful concept should not be discarded [and comments and reply]. Current Anthropology **40** (1999)
13. O'Reilly, C.: Corporations, culture, and commitment: Motivation and social control in organizations. California Management Review **31**(4) (1989) 9–25
14. Mead, M., Métraux, R., eds.: The Study of Culture at a Distance. University of Chicago Press (1953)
15. McFate, M.: The military utility of understanding adversary culture. Joint Force Quarterly **38** (2005)
16. Bozeman, A.B.: Strategic Intelligence and Statecraft: Selected Essays. Washington, DC: Brassey's (1992)
17. Boyd, R., Richerson, P.J.: Culture and the Evolutionary Process. University of Chicago Press (1988)
18. Kroeber, A.L., Kluckhohn, C.: Culture: A critical review of concepts and definitions. Papers of the Peabody Museum **47** (1952) 643–656

19. Harrison, R.J., Carroll, G.R.: Culture and Demography in Organizations. Princeton University Press (2006)
20. Schein, E.H.: Organizational Culture and Leadership (Jossey-Bass Psychology Series). 2 edn. Jossey-Bass (1996)
21. Bailey, G., Peoples, J.: Introduction to Cultural Anthropology. Wadsworth Publishing (1998)
22. Keesing, R.M.: Cultural anthropology: A contemporary perspective. Holt, Rinehart and Winston (1981)
23. Tylor, E.B.: Primitive culture. London: J. Murray (1871)
24. Linton, R.: The study of man: An introduction. New York: D. Appleton-Century (1936)
25. Harris, M.: Culture, people, nature: An introduction to general anthropology. New York: Thomas Y. Crowell (1975)
26. Ferber, J.: Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence. Addison-Wesley Professional (1999)
27. Sengers, P.: Anti-Boxology: Agent Design in Cultural Context. PhD thesis, Carnegie Mellon University (1998)
28. Hirshman, B.R., Carley, K.M., Kowalchuck, M.J.: Specifying agents in construct. Technical Report CMU-ISRI-07-107, Carnegie Mellon University (2007)
29. Hirshman, B.R., Carley, K.M., Kowalchuck, M.J.: Loading networks in construct. Technical Report CMU-ISRI-07-116, Carnegie Mellon University (2007)
30. Hirshman, B.R., Birukou, A., Martin, M.K., Bigrigg, M.W., Carley, K.M.: The impact of educational interventions on real & stylized cities. Technical Report CMU-ISR-08-114, Carnegie Mellon University (2008)
31. Balzer, W., Tuomela, R.: Collective intentions and the maintenance of social practices. Autonomous Agents and Multi-Agent Systems **6**(1) (2003) 7–33

A Method for Generating Social Networks from Meeting Transcripts

David André Broniatowski

Engineering Systems Division, Massachusetts Institute of Technology, E38-548, 77 Massachusetts Ave., Cambridge, MA 02141, USA
{david}@mit.edu

Abstract. The outcome of an intercultural negotiation often depends on the affinity that each negotiator feels with his or her peers. We take as given that the process of gaining membership within a professional community or discipline is one of acculturation. Cross-disciplinary committee meetings, wherein multiple experts from different specialties must negotiate a joint decision, are therefore intercultural negotiations. This paper presents a computational methodology for generating social networks from transcripts of such committee meetings, using medical device approval meetings within the Food and Drug Administration (FDA) as a data source.

Keywords: Linguistic analysis, Bayesian inference, committee decision-making

1 Introduction

Research within the sociology of professions has shown that professional specialties may be thought of as separate cultures [1]. Interactions between these specialties are therefore a form of intercultural negotiation. Just as different national cultures might negotiate the boundaries between their respective spheres of geopolitical influence, different professions are constantly engaged in redefining the boundaries of their own expertise [2].

The outcome of an intercultural negotiation often depends on the affinity that each negotiator feels with his or her peers. Given that different groups utilize terminology specific to their culture, we can identify cultural affinity through linguistic analysis. This paper is therefore aimed at determining which affinities exist between negotiators through a computational analysis of their language. We present a methodology whereby the different affinities of members engaging in an intercultural negotiation might be computationally inferred from a meeting transcript. In particular, we study interactions between committees of technical experts.

In the committees that concern us in this paper, information must be aggregated from multiple expert specialists to ultimately make a decision regarding whether a particular medical device should be approved for market. The decision of what information is

important and how it should be interpreted is the subject of negotiation up until the time that each committee member casts a vote. In such situations, we may expect concerns about bias, conflict of interest and deep uncertainty. The fact that different experts hold different perspectives and values makes it more likely that additional aspects of a problem will come under consideration. Nevertheless, this also makes it difficult to generate consensus on the interpretation of data. Experts' interpretations of data are likely influenced by institutional culture (e.g., a particular profession, specialty, or organization of which that expert is a part).

This work is aimed at developing a deeper understanding of how differential perceptions, e.g., due to different training or demographic features, impact upon multi-actor decision-making. Such differences are identified and isolated through an analysis of the language used by each speaker in the negotiation. This approach is based on the idea that language, beyond simply conveying information, is also an expression of identity. Language, as a system of abstraction, will necessarily reflect the speaker's direction of attention. More specifically, one's choice of terminology will tend to reflect those points of the negotiation that one finds salient. We utilize an empirical quantitative methodology based upon a computational linguistic analysis of meeting transcripts. Such a methodology can be extended to similar studies in other domains of interest to students of intercultural negotiation.

We use a modification of the Author-Topic (AT) model [3], a Bayesian inference tool used in the field of information retrieval, to discover linguistic affinity between committee members. We find that the resulting output may be used to construct social networks, analysis of which shows that committee members often group together by medical specialty and voting pattern. Such analyses might be used to uncover cultural affinity from transcripts where demographic and voting data are not available.

2 Why the FDA?

The empirical analysis mentioned above requires data in the form of committee meeting transcripts. These are often not recorded in textual form, or are proprietary to the organization that commissioned the committee. We therefore turn to transcripts of expert committees that are a matter of public record. The ideal data source must have the following attributes:

1. Must involve a negotiation (e.g., analysis or evaluation of a technological artifact).
2. Participation of representatives from multiple cultures (e.g., multiple experts from different fields or areas of specialization).
3. A set of expressed preferences per meeting (such as a voting record).
4. Multiple meetings, so as to enable statistical significance.

These requirements are met by the Food and Drug Administration's medical device advisory panels. The most uncertain, and therefore difficult, medical devices are reviewed by expert advisory panels prior to their exposure to the American market. A device's approval and future market diffusion often rests upon the panel's assessment of the device's safety. These panels are aimed at producing a recommendation, informed by the expertise and knowledge of panel members, which can supplement the FDA's "in-house" decision process. Multiple experts are consulted so that the group decision's efficacy can take advantage of many different institutional perspectives. Panel members' values and institutional contexts may differ, leading to different readings of the evidence, and therefore different recommendations. In health care, Gelijns et al. note that strictly evidence-based decisions are often not possible for the following reasons [4]:

1. A given data-set may be interpreted differently by different experts, especially in the presence of high uncertainty. Unless these experts can learn from one another, good decision-making might be impaired.
2. Patterns of technological change are difficult to predict, particularly when innovations are ultimately used for different purposes than originally intended (e.g., the off-label use of a drug or device approved by FDA or the use of the military's GPS system for civilian geo-location).
3. Even in the case of clear evidence, decision-makers may disagree on its implications due to differing value systems.

This suggests that a device's determination as safe or efficacious depends strongly on factors that are not within the purview of strictly "evidence-based" decision-making. Douglas and Wildavsky argue that these are largely shaped by the perceptions, and hence, the knowledge and expertise, of risk assessors [5]. Institutions that might impact decision-making include membership in a particular profession, specialty, or bureaucratic organization [4, 6-8].

3 How to Analyze Institutional Factors?

The approach taken within this paper is inspired by work within the anthropology and Science, Technology and Society (STS) literatures. In particular, the penetrating analyses of Mary Douglas note that social group membership affects perception of data [9]. By institution, we mean a set of social norms to which a particular community adheres. Institutional membership is conferred upon those who structure categories of causality in a manner that is consistent with institutional norms [7]. This is reflective of a wider principle that different cultures will selectively direct individuals' attention to the elements of a negotiation that are salient within their institutional structures. This is reflected in the fact that each culture, and indeed, each specialty, possesses its own unique language and jargon, which carries with it an implicit scheme for categorizing perceived

phenomena. Within the evolutionary economics literature, Nelson also notes the importance of written and oral language as a means of encapsulating and transferring tacit knowledge [10]. On the other hand, an outsider to the institution, who categorizes the world differently, may be unable to understand the discourse. This is because the specific jargon refers to commonly held sensory and social experiences that a member of another institution is unlikely to have directly encountered. The STS literature extends this notion by noting that language is used as a cognitive mechanism to delineate professional boundaries. The attention of experts within a specialty is therefore directed toward a given interpretation of a problem that is consistent with that expert's training, while simultaneously directing that attention away from other possible interpretations [11-14]. Casting "organization [as] the mobilization of bias", Cobb and Elder [15] recognize institution-specific symbolism in language, noting that the choice of terminology in defining a problem may be seen as a means of mobilizing support. Furthermore, the linguistic definition of a problem dictates, to some extent, its solution. Choosing to use more specialized technical words serves to narrow the range of subjective meaning of otherwise ambiguous terminology (such as "safety" or "efficacy" in FDA's context) to the specific world view of the users of that jargon. This implicitly redefines the problem according to a given speaker's particular interest. The same cultural institutions that drive selective perception and word choice may also be expected to confer a sense of identity. Thus, we may expect individual preferences to be correlated with institutional membership [16]. This motivates an analysis of language in order to be able to examine institutional factors in group negotiation.

The work cited above suggests that the determination of institutional and other interpersonal affinity might be identified through the use of common language and jargon. If so, we might conclude that they find the same elements of the problem salient. This further suggests cultural affinity. The most direct way of operationalizing these insights is to attempt to cluster speakers by the co-occurrence patterns of their discourses. In particular, we use Bayesian modeling, to determine whether actors within a committee meeting are using similar terminology to discuss the common problem to be solved.

The choice of Bayesian topic models is driven by a desire to make the assumptions underlying this analysis minimal and transparent while maintaining an acceptable level of resolution of patterns inferred from the transcript data. An earlier iteration of this work, based on [17], used Latent Semantic Analysis (LSA) [18], a simpler predecessor of topic models, to study the same corpus of FDA transcripts. In practice, LSA can identify and separate major sources of variance in word choice within a discourse, as in evaluating the divergence between two groups of speakers (e.g., identifying device sponsors versus committee-members within the resulting latent semantic space). LSA has some well-known limitations that stem from its use of Singular Value Decomposition (SVD) to analyze word co-occurrence patterns. Among these is the assumption that words are embedded within a Euclidean "semantic-space". This particular assumption breaks down when comparing words that are polysemous – i.e., having the same spelling but different meanings (compare "bat" the animal vs. "bat" in the context of baseball). LSA represents the location of these words in the Euclidean semantic space as the average over the two

separate locations – an incorrect representation. Furthermore, LSA assumes that the noise around each word's location in the Euclidean space is normally-distributed, an assumption that introduces increasingly more distortion into the analysis as a given speaker uses fewer words. These limitations make it difficult to resolve the linguistic attributes of individual speakers, particularly in the absence of extensive speaker data within a given meeting. Furthermore, the latent dimensions of the LSA feature space, which nominally correspond to latent concepts of a discourse, are often difficult to interpret. These limitations motivate the use of a Bayesian model (for an excellent comparison of LSA to Bayesian models of text analysis, see [19]).

Given our assumption that each speaker possesses a cultural signature in his or her word choice, we would like to have the identity of the speaker inform the selection of topics. We therefore use a variant of Rosen-Zvi et al.'s Author-Topic (AT) Model [3], which creates probabilistic pressure to assign each author to a specific topic. Shared topics are therefore more likely to represent common jargon. Like LSA, AT also uses a term-document matrix as input. The AT model is implemented as follows:

3.1 AT Model Structure and Implementation

We begin by parsing a committee-meeting transcript into a word-document matrix. Consider a corpus of documents, $\mathcal{D}$, containing n documents $d_1 \ldots d_n$. Consider, as well, the union of all words over all documents:

$$W = \bigcup_{i=1}^{n} d_n \qquad (1)$$

Suppose there are m>n words in W, $w_1 \ldots w_m$. We may therefore construct a "word-document matrix", **Y**, with dimensions m x n, where each element in the matrix, y_{jk}, consists of a frequency count of the number of times word j appears in document k. For the analyses reported in this paper, a word-document matrix was constructed using the Python 2.5 programming language. Non-content-bearing "function words", such as "is", "a", "the", etc., were pre-identified and removed automatically. In addition, words were reduced to their roots using PyStemmer, a Python implementation of Porter's Snowball algorithm [20]. The resulting corpus generally consists of ~25,000 word tokens, representing about m = 2500 unique words in about n = 1200 utterances.

3.2 AT Model Structure and Implementation

The Author-Topic model provides an analysis that is guided by the authorship data of the documents in the corpus, in addition to the word co-occurrence data used by LSA and all topic models. Each author (speaker in the discourse) is modeled as a multinomial distribution over a fixed number of topics that is pre-set by the modeler. Each topic is, in turn, modeled as a multinomial distribution over words. A plate-notation representation of

the generative process underlying the Author-Topic model is found in Fig 1. The Author-Topic model is populated using a Markov-Chain Monte Carlo Algorithm that is designed to converge to the posterior distribution of words over topics and authors that best matches the data. Information about individuals authors is included in the Bayesian inference mechanism, such that each word is assigned to a topic with probability proportional to the number of words by that author already in that topic, and to the number of times instances of that specific word appears in that topic. Thus, if two authors use the same word in two different senses, AT will account for this polysemy. Details of the MCMC algorithm implementation are given in [3]. The AT model was implemented in MATLAB by the author, based on the Topic Modeling Toolbox algorithm [21].

3.3 Model Parameters

The AT model requires the selection of two parameters. Each author's topic distribution is modeled as having been drawn from a uniform Dirichlet distribution, with parameter α. A Dirichlet distribution is used because it is the conjugate prior of the multinomial distribution. One may think of α as a smoothing parameter. Values of α that are smaller than unity will tend to more closely fit the author-specific topic distribution to observed data. If α is too small, one runs the risk of overfitting. Similarly, values of α greater than unity tend to bring author-specific topic distributions closer to uniformity. A value of α=50/(# topics) was used for the results presented in this paper, based upon the values suggested by [21] after extensive empirical testing. Similar to α is the second Dirichlet parameter, β, from which the topic-specific word distributions are drawn. β values that are large tend to induce very broad topics with much overlap, whereas smaller values of β induce topics which are specific to small numbers of words. Following the empirical guidelines set forth by Griffiths and Steyvers [21], and empirical testing performed by the authors, we set the value of β = 200/(# words).

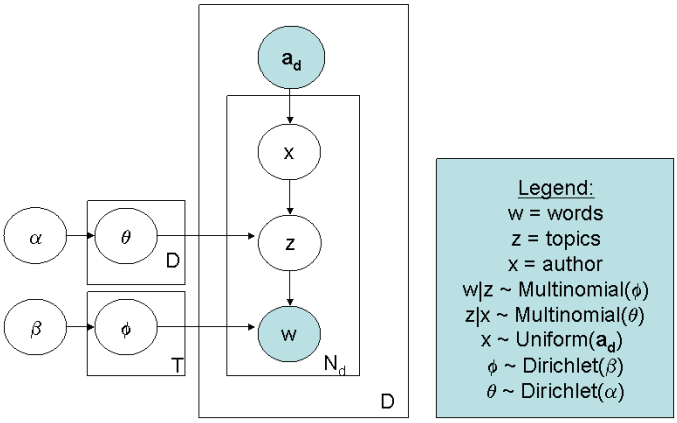

Fig. 1. A plate notation representation of the Author-Topic model from [3]. Authors are represented by a multinomial distribution over topics, which are in turn represented by a multinomial distribution over all words in the corpus.

3.4 Committee Filtering

It is often difficult to differentiate between voting panel members, particularly when they compared to all other meeting attendees (including device sponsor representatives, FDA representatives, non-voting members, etc). This is partially because they share procedural language in common. Therefore, a large proportion of the words spoken by each committee member would all be assigned to the same topic. This problem is solved using the AT model by creating a "false author" named "committee". Prior to running the AT model's algorithm, all committee voting members' statements are labeled with two possible authors – the actual speaker and "committee". Since the AT model's MCMC algorithm randomizes over all possible authors, words that are held in common to all committee members are more likely to be assigned to "committee", whereas words that are unique to each speaker tend to be assigned to that individual speaker. In practice, this allows individual committee members' unique topic profiles to be identified, as demonstrated below. In the unlikely case where all committee members' language is common, half of all words will be assigned to "committee" and the other half will be assigned at random to the individual speakers.

3.5 Sample Output

Table 1 displays the top five most probable word stems for each topic from a sample output run of the Author-Topic model applied to the FDA Meeting held on March 4th, 2002. Each topic may be identified by its most probable words:

Table 1. The top five word-stems for one run of the AT model on the corpus for the Circulatory Systems Devices Panel Meeting of March 4, 2002.

Topic Number	Top Five Word-Stems
1	'clinic endpoint efficaci comment base'
2	'trial insync icd studi was'
3	'was were sponsor just question'
4	'patient heart group were failur'
5	'devic panel pleas approv recommend'
6	'think would patient question don'
7	'dr condit vote data panel'
8	'effect just trial look would'
9	'lead implant complic ventricular event'
10	'patient pace lead were devic'

Within a clinical trial administered by the FDA, a device manufacturer must meet a certain set of clinical "endpoints", often manifested as a proportion of a population that is

free from disease or adverse events (e.g., device failure). Such trials typically have different endpoints for device safety and efficacy, both of which must be met. From this table, we can see that Dr. Konstam's major topic of interest involved questions of what was the appropriate clinical endpoint for the study in question (often a common debate on these panel meetings). It seems that he was particularly interested in the efficacy endpoints (as opposed to the safety endpoints). The AT model therefore shows some preliminary potential for identifying and summarizing the issues in which each speaker is interested.

4 Generation of Social Networks

The above methodology can give us insight into the topics of interest for each speaker. Nevertheless, topics, on their own, provide little direct information about how individual speakers might relate to one another. Instead, we would like to use the topic information provided by the AT model to generate a social network.

4.1 Network Construction

We would like to link together speakers who commonly use the same topics of discourse. In particular, we examine each author-pair's joint probability of speaking about the same topic.

$$P(X_1 \cap X_2) = \frac{\sum_{i}^{T} P(Z = z_i \mid X_1) * P(Z = z_i \mid X_2)}{\sum_{i,j}^{T,T} P(Z = z_i \mid X_1) * P(Z = z_j \mid X_2)} \quad (2)$$

We would like to be able to construct an Author-Author matrix, **A**, with entries equal to 1 for each linked author pair, and entries equal to 0 otherwise. This may be interpreted as a social network [22].

4.2 Author-Author Matrix Determination

The AT model outputs an Author-Topic matrix that gives the total number of words assigned to each topic for each author. This information must be reduced to the **A** matrix identified above. Given no prior information about a given author's topic distribution, we might assume that that such a distribution is uniform over all topics. Therefore, we might expect *a priori* that the joint probability that any author pair would be linked would be uniform. In other words, if there are 10 topics, we would expect every author-pair to have

a 10% probability of being linked, *a priori.* We consider an author pair to be linked within a given model iteration if that pair's joint probability exceeds what we would expect under a uniform distribution.

Each social network generated using this scheme is the result of one MCMC iteration. Multiple iterations, when taken together, form a probability distribution over a set of possible Author-Topic assignments, and therefore, feasible network topologies. We can expect that different iterations of the MCMC algorithm will yield different graphs.

Averaging over multiple MCMC iterations enables a social network to be created with weighted links, where the weight of each link is proportional to its frequency of occurrence among iterations. Nevertheless, the variability among draws from the MCMC algorithm suggests that links should not be weighted. Histograms of the distribution of these link frequency values tend to show a bimodal structure (see Fig. 3), suggesting that a description of author pairs as either connected or not connected is parsimonious.

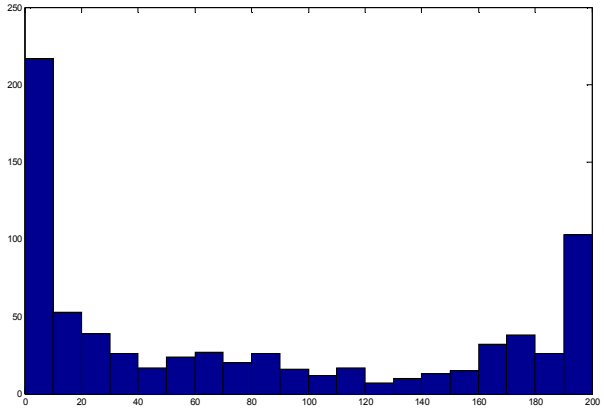

Fig. 2. Sample histogram of linkage frequency for an FDA Advisory Panel meeting. The horizontal axis is the link weight (i.e., the frequency with which author-pairs are connected over 200 samples from the AT model). The vertical axis is the link frequency of links with the weight specified by the abcissa (i.e., the number of author-pairs that are connected with the frequency specified by the abcissa). Note the existence of two modes located at the extremes of the distribution. This histogram defines a beta distribution.

If we consider each author-pair to have a fixed, binomial probability of being linked within a given meeting, then the histogram shown in Fig. 3 defines a beta distribution unique to that meeting. The parameters of this beta distribution can further serve as a metric of the extent to which a given meeting is polarized.

In order to generate the networks shown in this paper, all author-pairs that were linked in more than 95% of all iterations were considered strongly-linked. All author-pairs that were linked in more than 90% of all iterations were considered weakly-linked.

5 Preliminary Network Analyses

The previous section demonstrated how social networks can be built. The following section begins a preliminary analysis of the capabilities of the methodology outlined in this paper.

5.1 Grouping by Medical Specialty

The results of the above analysis methodology support the assertion that language and medical specialty are correlated. Nevertheless, some meetings display voting along institutional lines more clearly than do others. For example, Fig. 3 and Fig. 4 show a strong grouping by medical specialty.

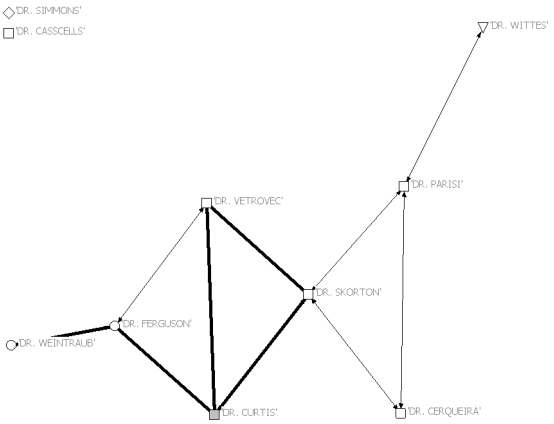

Fig. 3. Graph of the FDA Circulatory Systems Advisory Panel meeting held on April 24, 1998. This meeting yielded a consensus approval of the medical device under analysis. Node shape represents medical specialty (circles are surgeons, squares are cardiologists, diamonds are electrophysiologists). Dr. Curtis, in grey, was the committee chair. Dr. Casscells, the unlinked cardiologists, was the member of the committee who had most recently graduated from medical school.

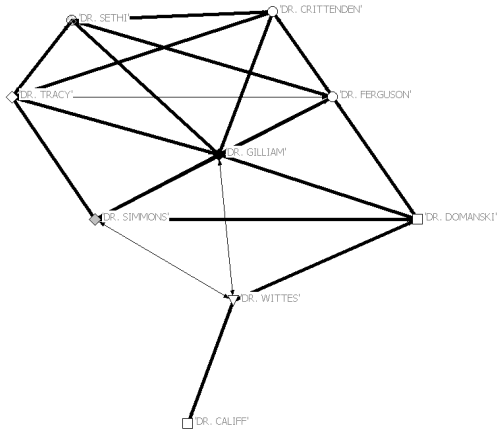

Fig. 4. Graph of the FDA Circulatory Systems Advisory Panel meeting held on October 27, 1998. This meeting yielded an approval of the medical device under analysis, with only Dr. Gilliam's dissent (in black). Node shape represents medical specialty (circles are surgeons, squares are cardiologists, diamonds are electrophysiologists). Dr. Simmons, in grey, was the committee chair. Dr. Sethi, the cross-hatched surgeon, was not present for the vote.

5.2 Grouping by Votes

In situations where the panel's vote is split, the method described in this paper can often isolate voting cliques (see Fig. 5 and Fig. 6). In some meetings, medical specialty and vote are aligned. This is the case in Fig. 5. In this meeting, all surgeons voted against device approval, whereas most cardiologists voted in favor. Radiologists' votes were split evenly between the two. In others, such as Fig. 6, there is a weaker correspondence. Nevertheless, all graphs show members of the same voting coalition to be connected. This

suggests that the device reviewed in these meetings might be evoking an identity that transcends medical specialty.

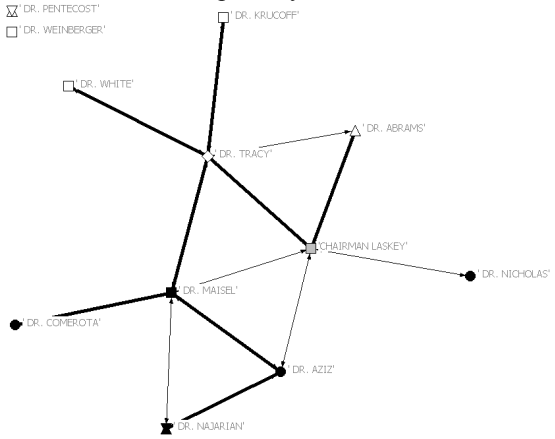

Fig. 5. Graph of the FDA Circulatory Systems Advisory Panel meeting held on April 21, 2004. This meeting yielded an approval of the medical device under analysis, although the panel was split (white, in favor; black against). Node shape represents medical specialty (circles are surgeons, squares are cardiologists, diamonds are electrophysiologists, triangles are neurologists, and hourglasses are radiologists). Dr. Laskey, in grey, was the committee chair.

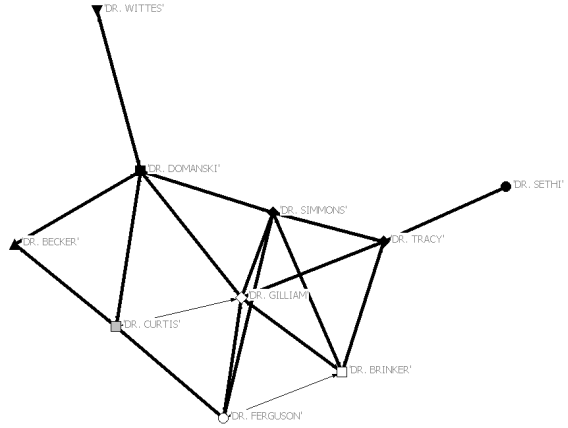

Fig. 6. Graph of the FDA Circulatory Systems Advisory Panel meeting held on June 6, 1998. This device was not approved. Node shape represents medical specialty (circles are surgeons, squares are cardiologists, diamonds are electrophysiologists, down triangles are statisticians, and up triangles are neurologists). Dr. Curtis, in grey, was the committee chair. Non-approval votes are in black; approval votes are in white. In this meeting, vote is not correlated with medical specialty.

In 30 of the 37 cases (~81%) for which graphs were generated, connectivity patterns could be explained using vote or specialty information alone. Many of the remaining seven cases may be explained using alternate notions of identity, such as race, gender or affiliation with a particular training institution (e.g., medical school).

5.3 The Impact of Ideology

On June 23, 2005 the Circulatory Systems Devices Panel held a meeting to determine whether a particular device should be approved for a Humanitarian Device Exemption. Such a meeting likely appeals to a sense of personal ethical responsibility that transcends medical specialty. In situations such as these, we might expect that individual votes and connectivity patterns will be more idiosyncratic and exhibit less coherence. Fig. 12 shows the connectivity pattern for this meeting. Note that this graph cannot be as easily partitioned by vote or by medical specialty, although yes voters do tend to congregate around the committee chair, whereas no-voters tend to congregate with the abstaining voters in the cluster near the upper-left. This suggests the presence of an idiosyncratic identity trigger.

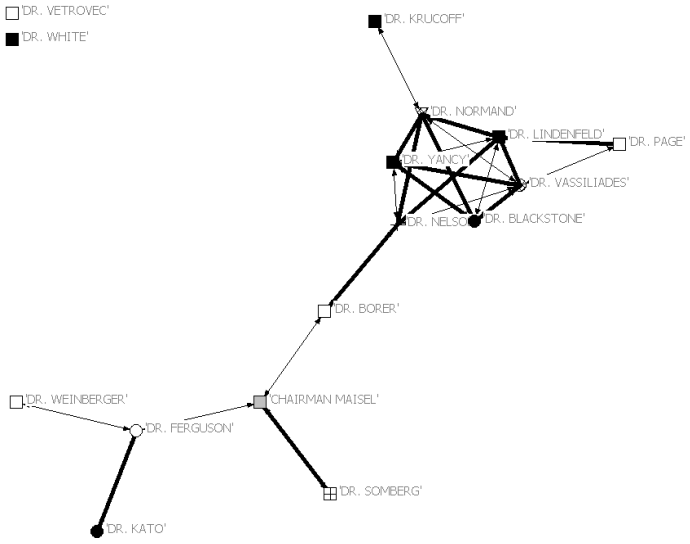

Fig. 7. Graph of the FDA Circulatory Systems Advisory Panel meeting held on June 23, 2005. Node color represents the vote (black is against humanitarian device exemption, white is in favor of humanitarian device exemption, cross-hatched is abstention. Dr. Maisel, the committee Chairman, is grey). Node shape represents medical specialty (circles are surgeons, squares are cardiologists, crosshairs are a pharmacologist, down triangle is a statistician, plus is a pediatric bioethicist).

6 Methodological Limitations and Future Work

The method outlined in this paper relies on meeting transcripts to generate empirical findings regarding committee decision-making. It is seemingly sensitive to the limitation

that not all committee members might express their views truthfully. Nevertheless, it is very difficult for individuals to avoid using jargon that they are familiar with. This is because word choice reflects identity, as shaped by cultural adherence to a social institution.

Perhaps a larger concern is the inability to differentiate between agreement and argument. Two actors are linked if they discuss the same topics. They may do so because they agree on some aspect of the device review, or because they are debating over interpretation of a given element of the debate. This is evident in the figures shown above, where linkage does not always indicate voting similarity. This reflects the notion that cultural similarity is not necessarily equivalent to similarity of preference. A major area for future research is the determination of valence on each of the links in the graph. Determining signs for these graph links will enable a more direct comparison of the voting record to the graphical structure. This, in turn, would enable the analysis of reputation effects and strategic voting – e.g., why don't people vote the way they say they will? Existing work on sentiment classification (e.g., [23]) might be applicable to this problem.

This research is aimed at the development of a quantitative methodology that may be applied to analyze multi-actor decision-making by committees of technical experts, an example of an intercultural negotiation. The methodology presented in this paper, although still preliminary, has been used to generate meaningful social networks from transcripts of FDA medical device advisory panel meetings. Future work will focus on applying this method to a larger number of cases with the intention of producing generalizable findings and developing theory.

7 References

1. Trice, H. M. *Occupational subcultures in the workplace*. Cornell University Press, 1993.
2. Collins, H., and Evans. R., *Rethinking Expertise*. University Of Chicago Press, 2009.
3. Rosen-Zvi, M., et al., "The author-topic model for authors and documents," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence* (Banff, Canada: AUAI Press, 2004), 487-494,
4. Gelijns, A. C., et al., "Evidence, Politics, And Technological Change," *Health Affairs* 24, no. 1 (2005): 29-40.
5. Douglas, M., and Wildavsky, A., *Risk and Culture* (Berkeley, CA: University of California Press, 1982).
6. Freiman, M.P., "The Rate of Adoption of New Procedures among Physicians: The Impact of Specialty and Practice Characteristics," *Medical Care* 23, no. 8 (1985): 939-945.
7. Brown, J. A. "Professional language: words that succeed." *Radical History Review* 34 (1986): 33-51.
8. Savage, D.A., "Professional Sovereignty Revisited: The Network Transformation of American Medicine?," *Journal of Health Politics, Policy and Law* 29, no. 4-5 (2004): 611-677.
9. Douglas, M., *How Institutions Think* (Syracuse, New York: Syracuse University Press, 1986).

10. Nelson, R.R., "On the uneven evolution of human know-how," *Research Policy* 32, no. 6 (June 2003).
11. Mulkay, M., Pinch, T., and Ashmore, M., "Colonizing the Mind: Dilemmas in the Application of Social Science," *Social Studies of Science* 17, no. 2 (1987): 231.
12. Rapp, R., "Extra chromosomes and blue tulips: medico-familial interpretations," *Living and Working with the New Medical Technologies: Intersections of Inquiry* (2000).
13. Winner, L., "On not hitting the tar-baby," *The Whale and the Reactor: A Search for Limits in an Age of High Technology. Chicago* 111 (1986): 138-154.
14. Cohn, C., "Sex and Death in the Rational World of Defense Intellectuals," *Signs* 12, no. 4 (1987): 718, 687.
15. Cobb, R.W., and Elder, C.D., *Participation in American Politics: The Dynamics of Agenda-Building* (Baltimore and London: The Johns Hopkins University Press, 1983).
16. March, James G. *A Primer on Decision Making: How Decisions Happen*. New York, NY: The Free Press, 1994.
17. Dong, A., "The latent semantic approach to studying design team communication," *Design Studies* 26 (2005): 445-461.
18. Deerwester, S., et al., "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science* 41, no. 6 (1990): 391-407.
19. Griffiths, T.L., Steyvers, M., and Tenenbaum, J.B., "Topics in Semantic Representation," *PSYCHOLOGICAL REVIEW-NEW YORK*- 114, no. 2 (2007): 211.
20. Porter, M.F., *Snowball: A language for stemming algorithms. October 2001*.
21. Griffiths, T.L., and Steyvers, M., "Finding scientific topics," *Proceedings of the National Academy of Sciences of the United States of America* 101, no. Suppl 1 (April 6, 2004).
22. Wasserman, S., and Faust, K., *Social Network Analysis: Methods and Applications*, 1st ed. (Cambridge University Press, 1994).
23. Thomas, M., Pang, B., and Lee, L., "Get out the vote: Determining support or opposition from Congressional floor-debate transcripts," in (Sydney, Australia: Association for Computational Linguistics, 2006), 327-335, http://www.aclweb.org/anthology/W/W06/W06-1639.

Acknowledgements
The author would like to thank Prof. Chris Magee, Prof. Maria Yang, and Dr. Joseph Coughlin for their extensive discussions and valuable mentorship, and the MIT-Portugal Program for its generous support.

A Dynamical Tool to Study the Cultural Context of Conflict Escalation

Lan Bui-Wrzosinska[1,2], Michele Gelfand[3], Andrzej Nowak[1,4] , Laura Severance[3], Urszula Strawinska[2,4], Magda Formanowicz[2] and Aleksandra Cichocka[1],

[1]University of Warsaw, Krakowskie Przedmieście 26/28, 00-927 Warsaw, Poland
[2]Warsaw School of Social Psychology, Chodakowska Street 19/31, PL - 03815 Warsaw, Poland
[3]University of Maryland, 1147 Biology-Psychology Bldg., College Park, MD 20742
[4]Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431

Abstract. The present article describes research in progress which is developing a simple, replicable methodology aimed at identifying the regularities and specificity of human behavior in conflict escalation and de-escalation processes. These research efforts will ultimately be used to study conflict dynamics across cultures. The experimental data collected through this methodology, together with case-studies, and aggregated, time-series macro data are key for identifying relevant parameters, systems' properties, and micro-mechanisms defining the behavior of naturally occurring conflict escalation and de-escalation dynamics. This, in turn, is critical for the development of realistic, empirically supported computational models. The article outlines the theoretical assumptions of *Dynamical Systems Theory* with regard to conflict dynamics, with an emphasis on the process of conflict escalation and de-escalation. Next, work on a methodology for the empirical study of escalation processes from a DST perspective is outlined. Specifically, the development of a progressive scenario methodology designed to map escalation sequences, together with an example of a preliminary study based on the proposed research paradigm, is presented. Implications of the approach for the study of culture are discussed.

Keywords: Conflict Escalation, Dynamical Processes, Cross-Cultural Differences

1 Introduction

On March 1, 2001 Afghanistan's puritanical Taliban Islamic militia began the destruction of statues across the country, including the almost 2,000-year-old world's tallest statue of Buddha in Bamiyan. Although the destruction of this invaluable evidence of human culture and civilization occurred across merely a few days, the damage done cannot be possibly compensated, and is practically irreversible.

On September, 11, 2001, two airplanes hijacked by al Qaeda suicide bombers crashed into the twin World Trade Center towers in New York, causing the death of 2,974 people, targeting one of the most powerful symbols of the Western world and destabilizing the global political, cultural, and economical status quo.

On September 30th, 2005, twelve caricatures of Mahomet were published by the Danish newspaper Jyllands-Posten under the title "faces of Mahomet." This publication instigated a series of conflicts reaching far beyond the initial

provocation, including death threats toward the authors, high level political repercussions and rising social unrest.

Understanding, predicting, and managing conflict are arguably among the most important challenges facing mankind. With increasing interdependence, the well-being of societies and their potential for growth and cooperation primarily depends on the way the global community is able to handle existing as well as emerging social conflicts. As the above examples attest, however, conflicts traverse cultures and are constantly changing and evolving. Moreover, culture and conflict are intricately related. Individuals initiate conflicts, which become entrenched and affect the culture in which they are embedded; and at the same time, individuals are themselves conditioned and influenced by culture in the types of triggers that initiate conflicts and the factors that affect their escalation and de-escalation. This makes conflict and culture both *local* – that is, deeply anchored in human experiences and actions, and *global* – with large-scale, system-level consequences. The multilevel dynamic character of today's conflicts presents a challenge to all social sciences, as it requires new tools to help understand, predict, and manage the constantly evolving, dynamical character of the phenomenon.

Recently, a paradigm widely used in other areas of science – the dynamical systems approach (DST)- has been applied to the study of social conflict [1], [2], [3]. This theoretical advance has opened new avenues for the study of complex systems of conflict, bringing computational models, computer simulations and advanced conceptual tools to bear on studying conflict (see: [4], [5], [6]). Computational models and simulations hold the potential to advance not only the understanding of the dynamic interplay between culture and conflict, but also to have predictive value, which is critical for applications. However, empirical data is also crucial to ensure that computational models have relevance and predictive power. Empirical data should utilize multiple methods of data collection, including case studies analysis, statistical, time-series data, field work, and experimental psychological data allowing for causal inference.

The present article describes a work in progress on the latter effort—experimental psychological data—which is a part of several larger initiatives: the Dynamics of Conflict initiative [7], and the MURI initiative [8], aimed at bringing together advances from the application of dynamical systems to intractable conflicts, with the latest lines of research on the cultural context of conflict and cooperation. The main focus of the present project is the cultural context of conflict escalation and de-escalation dynamics. Specifically, we report our efforts toward the development of a simple, replicable methodology aimed at identifying the regularities and specificity of conflict escalation and de-escalation patterns which ultimately can be used to study conflict dynamics across cultures. We believe that experimental data collected through this kind of methodology can help identifying relevant cultural parameters, and their effects on naturally occurring conflict escalation and de-escalation processes. This, in turn, can facilitate the development of realistic, empirically supported computational models.

In what follows, we first outline the theoretical assumptions of DST with regard to conflict dynamics, with an emphasis on the process of conflict escalation and de-escalation. Next, the work in progress on a methodology for the empirical study of

escalation processes from a DST perspective will be outlined. Specifically, the development of a *progressive scenario methodology* designed to track escalation sequences, together with an example of a preliminary study based on the proposed research paradigm, will be reported. Implications of the approach and methodological tools for the study of culture are then discussed.

1.1 The Dynamical Systems Approach to Conflict

Along with the dynamical systems approach to social psychology [9] social phenomena can be described with the use of some core Dynamical Systems Theory concepts. The term *dynamical system* is generally used to describe numbers of interconnected elements that change and evolve over time. From this perspective, for example, boiling water is viewed as the current state of a system of interacting molecules, the brain as a system of interacting neurons, or the society as a system of interacting individuals. A dynamical system can generally be conceptualized as the state of its elements at a given time; a system's behavior as a sequence of such states. To describe sequences of states, we need to identify key variables and parameters capturing the evolving characteristics of the system [9]. Although key parameters such as temperature and density for instance, describing a systems of interacting molecules in a state of steaming water, ice, or vapor seem relatively simple to measure and identify, specifying key parameters for the description of human systems undergoing different phases of conflict escalation still poses an important challenge to social sciences. It requires not only empirically informed, qualitative understanding of the phenomenon of conflict, but foremost the translation of qualitative, stable, social psychological properties into quantitative, measurable dynamical variables. Moreover, efforts toward the identification and definition of key social, psychological, and cultural parameters are tantamount to the mapping of such a system's behavior over time.

In the present project, we map conflict escalation and de-escalation as a sequence of one party's reaction to another party's conflict provoking behavior. From this point of view, the trajectory of escalation is a sequence of measures of the participant's behavior in response to the step-by-step increase of aggressiveness of another party in conflict over time. The response is measured on a behavioral scale reflecting the level of conflict intensity of the response, from very low (conciliatory acts) to very high (physical aggression). What is interesting from the point of view of the present article, is that patterns of responses for the same scenario of conflict provocation by another party can vary across different social, cultural, and psychological conditions: people can escalate gradually in response to gradual intensification of aggressive behaviors from the other party, but the same conditions can also lead to exaggerated response or, conversely, resistance to change and stabilization at a given level of intensity. Responses trajectories can also progress along some repeated cycles, or have unpredictable, irregular character.

One way to formally portray and systematize such results is to describe the dynamical properties of conflict escalation trajectories as attractor's dynamics. Generally, the dynamical systems approach to social psychology [9] identifies and describes attractors in social systems as regions, toward which trajectories in a state

space converge with time. It is common to distinguish among four classes of attractors [10], [11]: fixed-point, periodic, quasiperiodic, and chaotic. Here, we concentrate on fixed-point attractors. The method we use here to assess attractor dynamics is to actively perturb the system through a sequence of conflict provocation stimuli. If a single, fixed-point attractor exists, the system will always return to the same state after some time, thus one party's response will return to the same level of conflict intensity, regardless of the influence from the other party. In the case of multiple fixed-points, small perturbations will result in the system returning to its original state, but further changes of the control parameter may result in the system moving toward a different equilibrium: threshold effects are to be expected in the responses patterns. In dynamical social psychological terms similar dynamics have been understood as catastrophic scenarios of change [12], and will be referred to as catastrophic (as opposed to gradual, incremental) escalation. Properties of such scenarios are of particular relevance for de-escalation and practical applications: the *hysteresis* effect described in catastrophe theory [13], for instance, explains how crossing certain thresholds in conflict escalation leads to irreversible changes, undermining the potential for further de-escalation. In our project, however, this would rather be a post hoc conceptualization of emerging properties and parameters explorations than mathematically supported, precise models.

Our goal is to explore the variance of people's response trajectory in different cultural, social, and psychological contexts. The work we are advancing here is aimed at identifying naturally occurring sequences in escalation dynamics, but at the same time controlling for cultural conditions which could emerge as critical parameters for escalation dynamics. We are thus not yet at the stage of empirical testing of existing models of conflict escalation, but rather at the preliminary stage of parameters identification, as well as exploration of dynamical properties, naturally occurring patterns, triggers and qualitative shifts in controlled, laboratory conditions. This stage, we believe, is critical for the further development of models that have social psychological relevance. Below we discuss the initial development of a tool – *the progressive scenario methodology* – a work in progress toward experimental data collection aimed at testing the role of cultural parameters on interpersonal conflict escalation and de-escalation dynamics.

2 Development of the *progressive scenario tool*

The *progressive scenario tool* is mapping the response of one party to another party's conflict provocation behavior. As a starting point, the main parameter describing the system's behavior is derived from one of the most robust theories in conflict theory, Deutsch's theory of conflict cooperation and competition [14]. Morton Deutsch's seminal question, "under what conditions will a conflict follow a constructive versus destructive path?" is investigated in a dynamical way, through the translation of the "destructiveness" variable into concrete behaviors ranked according to their level of destructiveness versus constructiveness. The manipulated parameter, representing the stimulus responsible for changes over time, is a linear progression of the other party's conflict provocation behavior. Through the use of this tool, numerous independent

variables can be manipulated in order to track their effect on the course of escalation / de-escalation. Below, we describe the two components of the tool: (1) the stimuli and (2) the response scale.

2.1 Stimuli

The stimuli consist of a series of short descriptions (vignettes) of gradually escalating and de-escalating conflict behaviors displayed by a colleague at work in a situation of task interdependence ("you are working on a common project at work"). Fourteen subsequent vignettes are scaled according to the level of destructiveness and aggressiveness of the behavior they represent: the first 7 scenes outline a scenario of progressive escalation of provocation by a colleague at work, from a relatively mild disagreement ("Your colleague criticizes your work") to open confrontation and humiliation ("During a company picnic, your colleague insults your partner / relative publicly"). The remaining 7 scenes outline a progressive de-escalation scenario, with descriptions of conciliatory behaviors aimed at reversing each escalatory step ("Your colleague apologizes publicly for his inappropriate behavior toward your partner / relative").

2.2 Response Scale

The response scale includes a list of 30 behaviors, scaled with regard to the level of destructiveness to the relationship (between the parties) that they represent, from relatively constructive ("talking it over") to extremely hostile and destructive behaviors ("hurting him / her as much as possible"). Items were generated via focus groups conducted with individuals working in organizations as well as discussions with subject matter experts (professional mediators, and scholars from the conflict resolution field). Subsequently, large samples of individuals scaled items along conflict dimensions by employing multidimensional scaling techniques. These efforts were aimed at collecting qualitatively informed items (focus groups and subject matter experts), with the possibility to translate qualitative properties to quantitative data (scaling of the items along social psychological dimensions), and thus map a party's response trajectory on the defined phase space with some relative precision given the qualitative character of the data. In sum, the response scale is designed to measure changes in the order parameter (destructiveness of the response behavior) of the system.

2.3 Conflict Trajectories

Results from the questionnaire (responses on the scale of possible behaviors for each level of the provocation) can be mapped as a trajectory on a two dimensional space. The space is defined by the level of conflict provocation displayed by the other party (control parameter), and by the level of response destructiveness and aggression (order parameter). The trajectory represents a sequence of states. Each state is

described as participant's "map of possible behaviors" (the ensemble of behaviors considered as possible to display at a given time) in response to a given level of conflict provocation from the other party. This allow for the identification of patterns of dominant behaviors, but also latent clusters of behavior that remain stable across situations. Results from studies using the described questionnaire constitute a starting point for modeling, causal inference, as well as for the testing of the effect of various parameters on a laboratory simulation of escalation dynamics.

3 Example from a Preliminary Study

As an example here, we describe a preliminary study investigating the impact of relational closeness between parties on conflict escalation and de-escalation trajectories. The study revealed that closeness induces abrupt changes and nonlinear trajectories in conflict, while more distant relationships are characterized by gradual escalation trajectories. Results from this study [15] demonstrate that close relationships induce trajectories displaying a major shift from a series of responses, where, despite contentious behaviors from the other party, the responses are stabilized at a very low level of destructiveness, until a threshold is passed. At this point, the trajectory follows a sudden shift to a sequence of responses characterized by extremely high levels of destructiveness and open aggression. This type of dynamics is well known in nonlinear physics, and thus could be better understood with the use of DST concepts. Empirical results show a nonlinear progression of responses from one stable state of incontestably positive relations toward another stable state [16]; this finds coherent explanation as attractors dynamics, from the "friendship attractor" toward the "enmity attractor". Conversely, more distant relationship were associated with more gradual escalation patterns, where mid-range levels of aggression provoked intermediate responses. It is important to note that single static measures at a given moment in time would not predict the paradoxical effects of such conditions on the system's dynamics: from a static point of view, close friends are expected to uphold a stable, conflict free relationship [17]. However, the DST perspective demonstrates that this is true, but only for low levels of the control parameter (level of provocation).

Further exploration of responses items revealed that dimensions other than destructiveness, as well as triggers for nonlinear dynamics could have emerged as control parameters for the escalation process. For example, trust appeared to be critical in close relationships, and thus in a situation of rupture of trust, a shift has occurred in close relationships conditions, while this factor appeared irrelevant in the distant relationship condition. Such shifts between control parameters are being further investigated, and open an interesting line of research for the study of cultural differences. Results from this study are a basis for further development of theoretical and computational models exhibiting and extrapolating dynamical properties emerging from laboratory simulations.

4 Culture and Conflict Escalation

Conflict is a universal phenomenon, yet the way in which culture affects conflicts can vary dramatically across cultures [18] [19]. Nevertheless, cross-cultural research on conflict dynamics is in its infancy. Much research examines static differences in *stable conflict styles* with little or no attention to the dynamics of conflict across cultures. Building on our prior work, we are now developing new tools and examining the impact of culture on conflict escalation. Several initiatives are underway for the study of culture and conflict escalation. Through in-depth interviews in Egypt, Iraq, Jordan, Lebanon, Pakistan, Turkey, and the UAE, we are extracting local conflict episodes to develop new stimuli and new behavioral responses that are generalizable to the Middle East.

With our new tools, we will examine how cultural factors that are relevant to the Middle East, the U.S., and Asia—and in particularly—honor, dignity and face, respectively affect conflict dynamics [20] [21]. Cultural logics of honor, face, and dignity imply different trajectories of aggression. For example, people in *honor cultures* have a "keen sensitivity to the experience of humiliation and shame, sensitivity manifested by the desire to be envied by others and the propensity to envy the successes of others" ([22, pp. 116]. In such cultures, individuals are expected to go to great lengths to uphold the reputation of oneself and one's family and to avoid appearing vulnerable [23]. Reputation is critical within cultures of honor, while payback serves as an organizing principle for individuals' interactions when they have been provoked ([21, 23, 24, 25, 26]. Individuals from cultures of honor are thereby expected to respond *quickly* and with high levels of destructiveness in order to show they are not vulnerable. Individuals are expected to have a quick reaction to even minor acts, due to the importance of reputation maintenance (the premium placed on having a 'tough' reputation necessitates a strong reaction to seemingly small breaches of respect). It is also predicted that individuals will continue to act aggressively towards the perpetrator even after time has passed since the initial transgression; that is, there will be little 'cooling off,' consistent with "hysteresis effects" described in DST [5]. Particular triggers such as damage to female honor, shame, and humiliation are expected to also be important control parameters affecting thresholds of escalation in honor cultures.

In contrast to honor, wherein self-worth can be taken away by other's actions, individuals in *dignity cultures* are theoretically born with equal worth and rights which cannot be taken away by others [21, 25]. In dignity cultures, external evaluations matter little, while internal valuations are of the utmost importance. Values such as autonomy, freedom, and standing up for one's beliefs play a crucial role in dignity cultures. Such cultures are also likely to endorse rationality, strong person-task separation, and an independent self-construal [27]. We hypothesize that individuals from dignity cultures will generally react to increasingly aggressive acts in a linear fashion. Put differently, the escalation of aggression may be described as a rational, tit-for-tat strategy. In the same manner, individuals should react with decreasing levels of aggression as a perpetrator attempts to de-escalate the situation by apologizing or attempting to restore the relationship. However, particular triggers are expected to result in more severe reactions on the part of a victim from a culture of dignity, such as insults towards one's genuineness and challenges to one's freedom, rights, and autonomy [28] or status as an equal member of society [25], and thus might be seen as critical control parameters of conflict trajectories in dignity cultures.

Finally, in *face cultures*, individuals place a large premium on external evaluation of the self, while lending little credence to internal evaluations. *Face* represents an individual's claimed sense of positive image in the context of social interaction [29]. Similar to honor cultures, upholding the reputation of both the individual and the family is critical. Face cultures tend to have strong norms for communal responsibility, person-task interdependence, and maintaining harmony.. Compared to honor and dignity cultures, we expect that individuals from face cultures will be slow to react to initial aggressive acts and will react with less destructiveness. Over time, however, with continued provocation, we expect that conflict dynamics in face cultures can take on a *"catastrophic escalation"* pattern. Moreover, certain transgressions may trigger strong aggressive reactions in face cultures such as public criticism or embarrassment, communal shame, or violations of duties [30].

5 Discussion

The dynamical-systems approach to the study of culture, negotiation, and collaboration offers the potential to enhance our understanding of the culture and conflict in three distinct ways: metaphorically, mathematically, and empirically. First, dynamical-systems theory offers a rich array of new metaphors, constructs, and principles which might be fruitfully applied to the culture and conflict literature. Dynamic system constructs such as attractors, emergence, and self organization can serve as useful metaphors to help the researcher understand the dynamic nature of conflict and culture. Second, the dynamical systems approach provides the social scientists tools facilitating the mathematical description of the hypothesized mechanisms underlying specific culture and conflict dynamics. Thus, although social science theory is typically expressed verbally, the dynamical systems tools translate these theories into computer simulations. This will allow identification of assumptions inherent in our theories, but difficult to identify when theories are maintained in their verbal form. Finally, the dynamical-systems approach has implications for the types of empirical methodologies developed and employed in research. Typically, traditional social sciences focus on the central tendency of variables and ignore important dynamics reflected in variables' variances. Further, dynamical-systems models and methods push the social sciences to focus on events as they unfold overtime to understand the general *pattern of interactions* of the parties overtime.

The dynamical tool described in this article is a work in progress. Current versions, periodically updated are made available on the following web platform:

http://houselab.eu/DEV/iccc

Acknowledgements

This research is based upon work supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-08-1-0144.

References

1. Coleman, P. T., Vallacher, R., Nowak, A., & Bui-Wrzosinska, L. (2007). Intractable conflict as an attractor: Presenting a dynamical model of conflict, escalation, and intractability. *American Behavioral Scientist, 50*,. 1454-1475
2. Coleman, P. T., Bui-Wrzosinska, L., Vallacher, R., & Nowak, A. (2006). Approaching Protracted Conflicts as Dynamical Systems: Guidelines and Methods for Intervention. In A. Schneider & C. Honeyman (Eds.), *The Negotiator's Fieldbook* (pp. 61-74). Chicago: American Bar Association Books
3. Vallacher, R., Nowak, A., Bui-Wrzosinska, & Coleman, P. T. (2006). Attracted to conflict: A dynamical perspective on malignant social relations. In A. Golec & K. Skarzynska (Eds.), *Understanding social change: Political psychology in Poland, (*pp. 33-50). Haauppague NY: Nova Science Publishers Ltd.
4. Guastello, S., Koopmans, M., & Pincus D. (Eds.), (in press). *Chaos and complexity in psychology: The theory of nonlinear dynamical systems.* Boston: Cambridge University Press.
5. Nowak, A. & Vallacher, R. R. (1998). *Dynamical social psychology.* New York: Guilford Press.
6. Vallacher, R. R. & Nowak, A. (2007). Dynamical social psychology: Finding order in the flow of human experience. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles (2nd* ed, pp. 734-758). New York: Guilford Publications.
7. Nowak, Coleman, Vallacher, Bui-Wrzosinska, Bartoli, Liebovitch, www.dynamicsofconflict.iccc.edu.pl
8. Gelfand, Project Interaction, www.muriculture.com
9. Nowak, A. & Lewenstein, M. (1994). Dynamical systems: A tool for social psychology? In Vallacher, R. R. & Nowak, A. (Eds.), Dynamical systems in social psychology (pp. 17-53). San Diego, CA: Academic Press.
10. Eckmann, J.P. and Ruelle, D. Ergodic theory of chaos and strange attractors. Review Mod. Physics, 57(3):617--656, 1985.;
11. Schuster (1984): "Deterministic Chaos" An Introduction, Monographie 248 Seiten, Physik Verlag, Weinheim
12. Nowak, A. & Vallacher, R. R. (1998). Dynamical social psychology. New York: Guilford Press.
13. Thom, R. 1975. Structural Stability and Morphogenesis. An Outline of a General Theory of Models. Reading (Massachusetts): W. A. Benjamin, Inc.
14. Deutsch, M. (1973). *The resolution of conflict: Constructive and destructive processes*. New Haven, Yale University Press.
15. Bui-Wrzosinska, L., Nowak, A., Cichocka, A., Coleman, P.T., Vallacher, R., & Tan, R.Y. (in press) *Escalation and de-escalation patterns in interpersonal conflict: A dynamical systems approach.*
16. Tesser, A. & Achee, J. (1994). Aggression, love, conformity and other social psychological catastrophes. In R. Vallacher & A. Nowak (Eds.) *Dynamical Systems in Social Psychology* (pp. 95-109). San Diego: Academic Press.
17. Parker, J.G., & Asher, S.R. (1989). Friendship Quality Questionnaire-Revised: Instrument and administration manual. University of Michigan.
18. Gelfand, M.J., Erez, M., Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology, 58*, 479-514.

19. Imai, L., & Gelfand, M.J. (In press). The culturally intelligent negotiator: The impact of cultural intelligence (CQ) on negotiation sequences and outcomes. *Organizational Behavior and Human Decision Processes*.
20. Nisbett, R.E.,& Cohen, D. (1996). *Culture of Honor: The Psychology of Violence in the South*. Denver, CO: Westview Press.
21. Leung, A.K., & Cohen D. (2009). Within and between-culture variation: Individual differences and the cultural logics of honor, face, and dignity cultures. Working paper.
22. Miller, W.I. (1993). *Humiliation: And other essays on honor, social discomfort, and violence.* Ithaca, NY : Cornell University Press.
23. Vandello, J.A., & Cohen, D. (2003). Male honor and female fidelity: Implicit cultural scripts that perpetuate domestic violence. *Journal of Personality and Social Psychology, 84,* 997-1010.
24. Delgado, A. R., Prieto, G., & Bond, R. A. (1997). The cultural factor in lay perceptions of jealousy as a motive for wife battery. *Journal of Applied Social Psychology, 27,* 1824–1841.
25. Kamir, O. (2006). Honor and dignity in the film *Forgiven*: Implications for sociolegal theory. *Law & Society Review, 40,* 193-234.
26. Rodriguez Mosquera, P. M., Manstead, A. S. R., & Fischer, A. H. (2000). The role of honor-related values in the elicitation, experience, and communication of pride, shame, and anger: Spain and the Netherlands compared. *Personality and Social Psychology Bulletin, 26,* 833–844.
27. Markus, H.R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224-253.
28. Mosquera, Manstead, & Fischer (2002). The role of honour concerns in emotional reactions to offences. *Cognition and Emotion, 16,* 143-163.
29. Oetzel, J.G., & Ting-Toomey, S. (2003). Face concerns in interpersonal conflict: A cross-cultural empirical test of the face negotiation theory. *Communication Research, 30,* 599-624.
30. Liao, Y. & Bond, M.H. (Under review). The dynamics of face loss following interpersonal harm.

A Game-Theoretic Approach to Modeling Cross-Cultural Negotiation

Miroslav Dudík and Geoffrey J. Gordon

Carnegie Mellon University, Machine Learning Department,
5000 Forbes Avenue, Pittsburgh, PA 15213,
{mdudik,ggordon}@cs.cmu.edu

Abstract. Faithful models of negotiation should capture aspects such as subjective incentives, imperfect information, and sequential interaction, while providing explanation for behaviors such as bluffing, trust building, and information revelation. All of these objectives are elegantly addressed by theory of sequential games, and some of these phenomena have no convincing explanation without game theory's key assumption, namely, that of the rationality (or approximate rationality) of the negotiators. In this paper we discuss a game-theoretic approach to modeling negotiation. In addition to accounting for a range of behavior and reasoning styles we also address several aspects specific to cross-cultural negotiation. We argue that the existence of culture-specific beliefs and strategies can be explained by the existence of multiple game-theoretic equilibria. Within a culture, repeated interaction and learning lead to an equilibrium. On the other hand, across cultures, infrequent interaction leads with high probability to disparate (and often incompatible) equilibria. We hypothesize that inefficiency in cross-cultural negotiation can be attributed to this incompatibility. We discuss recently-developed algorithms that can be used to fit models of culture-specific behavior from data while incorporating rationality constraints. We anticipate that the additional structure imposed by rationality constraints will yield both game theoretic insights and also result in statistical advantages.

1 Introduction

Negotiation among multiple parties, whether in the context of business agreements or in the context of resolution of political conflicts, is an important problem studied both by economists and psychologists [1, 2]. However, computational modeling efforts have been somewhat modest—the most relevant literature focusing on behavioral modeling of the economic interaction [3, 4]. Prevalent computational approaches model the opponents probabilistically, which fails to fully account for the fact that opponents are aware that we are optimizing our own objectives. This limitation is addressed by concepts of game-theoretic equilibria. In this paper we outline the challenges of modeling negotiation and show how game-theoretic treatment addresses them. We also show how game theory accounts for the rise of culture-dependent beliefs about others, and explains the inefficiency in cross-cultural negotiation.

By modeling a negotiation, we mean the effort to capture a wide range of observed *behaviors* as well as *reasoning styles* of individual agents—behaviors include actions such as "sharing information" (e.g., communicating one's preferences), whereas reasoning includes internal judgments such as "the other agent just offered a compromise, so she probably values fairness". Since both behavior and reasoning take place in concrete situations that arise over the course of multi-agent interaction, a successful modeling approach needs to account for the structure of this interaction. Two key aspects of the interaction are the *imperfect information*—for example agents initially know only their own objectives, but has only probabilistic expectations about the objectives of others—and *sequential decisions*, for example, speaking turns.

A partial solution to the previous challenges is provided by *partially observable Markov decision processes* (POMDPs) and similar latent-state sequential models [3, 4]. They take a view of a single agent who is trying to optimize her own utility function in a sequential interaction with imperfectly observed environment which includes all other agents. The key limitation of POMDPs is that they do not capture *strategic reasoning*. Strategic reasoning accounts for the fact that the environment includes other agents who are trying to optimize their own utility functions. While POMDPs can contain latent state variables describing opponents' beliefs and policies (an approach known as *opponent modeling*), a POMDP cannot account for strategic behavior such as randomization among moves. Randomization is essential to human-like behavior: e.g., a human would never play a deterministic strategy in rock-paper-scissors. In order to give a principled account of strategic behavior, we will therefore model negotiation as an extensive-form game. Extensive-form games capture both sequential structure of interaction and imperfect information while providing a wide range of behaviors and reasoning that arise solely as a result of the rationality of an agent.

In Section 2 we describe extensive-form games in more detail. However, in practice we usually work with their compact representations such as multi-agent influence diagrams (MAIDs), also described in Section 2. In Fig. 1, we give an example of a MAID model of negotiation. Each agent has a privately known objective, formally denoted as the type, which may include aspects such as collectivism or individualism. Knowing their types (but not the types of others), agents take turns talking to each other and at the end they make their final decision (take the final action). The payoff (or utility) that they derive from the outcome depends on their type, their final action and on the final action of others. Arcs in the graph denote these dependencies—for example, when the second agent speaks, she is aware of her type as well as of what the first agent just said. Some additional arcs are implied by the condition of the perfect recall (agents do not forget anything throughout the game)—for example, when the second agent speaks for the second time, she continues to be aware of her type and all that was said before. The purpose of the communication is to foreshadow the final action and achieve some coordination.

Besides modeling the process of negotiation, we are particularly interested in modeling negotiation among participants from different cultures. In particular,

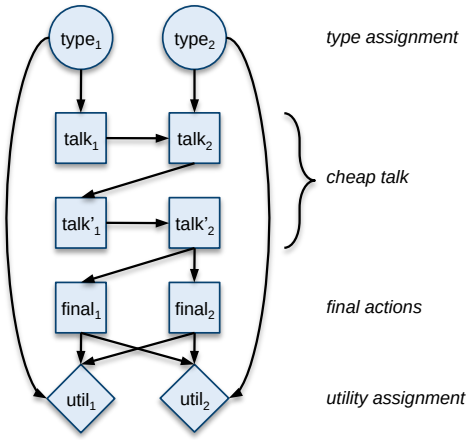

Fig. 1. An example game-theoretic model of negotiation between two agents.

we seek explanations for the inefficiency of cross-cultural negotiation. Using our game-theoretic setup, we hypothesize that this inefficiency can be attributed to the existence of multiple equilibria in negotiation games. One of the key problems in maintaining an equilibrium is the problem of believable and reliable information sharing. There are many possible ways for a culture to solve the information sharing problem. Within a culture, repeated interaction and learning lead to an equilibrium. On the other hand, infrequent interaction across cultures means that different cultures will likely converge to different (and often incompatible) equilibria.

While game theory addresses issues that cannot be addressed by opponent modeling, its assumptions are somewhat idealized. Specifically, perfect rationality and perfect recall are not accurate descriptions of real players. Yet, we believe that by capturing salient aspects of behavior game theory offers highly suitable models for analyzing negotiation. In the remainder of this paper we outline our modeling approach in more detail and discuss how game-theoretic equilibria yield various reasoning and behavior styles.

2 Game-theoretic Setup

We study a sequential game-theoretic formalism known as extensive-form games (EFGs) [5]. Let N be the number of agents, denoted as $n = 1, \ldots, N$. An extensive-form game is represented by a *game tree*, where non-leaf nodes represent either randomness (an action by nature) or an action by one of the agents. Non-leaf nodes of agents are partitioned into *information sets*. In each information set i a unique agent n is required to act. The agent knows only the identity of the information set, but cannot distinguish among the nodes in it, which represents partial information. Nodes in the same information set have the same set

of available actions. For each such action a, there is an edge labeled a leaving every node in the information set. The game begins in the root of the tree; agents and nature take turns until reaching a leaf. Each leaf contains an assignment of utilities to individual agents. The collection of information sets of an agent n is denoted $I(n)$. A deterministic behavior of the agent n is described by a *pure strategy*, which is a tuple $s_n = (s_i)_{i \in I(n)}$ with s_i specifying which action to take in each information set. The vector $s = (s_n)_{n \leq N}$ of pure strategies of all agents is referred to as the *strategy profile*. Nature's actions are drawn from some predetermined distributions.

As common in game theory, we restrict our attention to EFGs with *perfect recall*. Perfect recall means that agents do not forget any information over the course of the game. Formally, this requires that paths reaching nodes in an information set $i \in I(n)$ are indistinguishable by n, i.e., they contain identical sequences of n's information sets, and the agent n took identical actions in those information sets.

Instead of working directly with the game-tree representation, we work with a more succinct representation called *multi-agent influence diagrams* (MAIDs) [6]. Multiagent influence diagrams are game-theoretic generalizations of Bayes nets. Similarly to Bayes nets, MAIDs are represented by directed acyclic graphs. They have three types of nodes: (i) *decision nodes* (represented as rectangles), where a specified agent assigns a variable given the values of the parent variables, (ii) *chance nodes* (represented as ovals), where nature randomly assigns a variable conditioned on the values of the parent variables (according to a specified conditional distribution), (iii) *utility nodes* (represented as diamonds), where a specified agent receives utility as a function of the values of the parent variables; utility nodes have no outgoing edges. Any topological order specifies the order in which the play proceeds, for each node, distinct assignments of its parent variables correspond to distinct information sets.

To model negotiation, we consider MAIDs with a structure similar to the one in Fig. 1. First, nature assigns to each agent a *type* representing their utility function. As the arcs indicate, agents know their own types but not the types of others. Type assignments are followed by several rounds of *cheap talk*: "talk", because actions are fully observed, and "cheap", because actions do not directly result in any utility. After cheap talk, each agent carries out the *final action*. The combination of final actions of all agents determines the utility each agent receives.

Some strategic behavior, such as trust building, needs to be modeled in the context of a repeated interaction. In that case, the final action is followed by each agent's noisy observation of the final actions of other agents, and after which more rounds of cheap talk ensue followed by another final action, etc. We assume that the type assignment does not change between the stages.

The game-theoretic structure outlined so far captures only the structure of the interaction (sequential decisions and imperfect information). To capture the reasoning and especially strategic reasoning, we next shift our attention to equilibria in extensive-form games.

3 Equilibrium Reasoning

The most common equilibrium concept is that of Nash equilibrium. Unfortunately, many questions related to finding Nash equilibria are intractable, which we believe casts doubt on their suitability as behavioral models. Instead we will work with *extensive-form correlated equilibria* (EFCEs) [7], which can be computed more efficiently and, in addition, can be achieved as results of repeated interaction among learning agents [8]. This latter property lends support to the claim that EFCEs are a suitable model of negotiation behavior among same-culture agents. (We will return to cross-cultural negotiation below.)

Formally, an EFCE is a probability distribution over strategy profiles implemented by a moderator as follows. As information sets are reached during the game, the moderator suggests an action privately to the corresponding agents. Agents can either follow the moderator's suggestion or deviate. When agents deviate, they stop receiving suggestions from the moderator and must follow their own strategies. The probability distribution $p(s)$ is an EFCE if none of the agents can benefit by deviating.

The equilibrium distribution $p(s)$ determines the equilibrium reasoning of agents. For example, if the game reaches the information set i, belonging to the agent n, and the moderator suggests to take the action a then the agent n can calculate the posterior distribution $p(s_{-n} \mid s_i = a)$ which corresponds to the belief about other agents. Under this belief, the agent cannot improve her payoff by deviating from the suggestion a. This type of behavior will indeed be optimal, conditioned on the other agents' strategies, provided that the equilibrium distribution is the common knowledge (and thus beliefs are mutually compatible).

4 Cross-cultural Negotiation

In general, there are many EFCEs, and because of limited cross-cultural interaction we hypothesize that different cultures converge to different EFCEs. Using the algorithm of [9], it is possible to find EFCEs consistent with given negotiation transcripts. Negotiation transcripts will typically contain some information about agent types (agent personalities or details of their role), cheap talk (coded speaking turns), and final actions (the outcome of the negotiation). Given a set of same-culture negotiation transcripts, we can try to fit an equilibrium that matches the observed frequencies of various events such as "providing information by collectivist agents at the beginning of the game", or "proposing a multi-issue offer within two rounds after another agent describes their priorities". If our hypothesis is correct, the inclusion of equilibrium constraints should yield more accurate estimates of reasoning patterns (measured as beliefs at different points in the game) compared with the predictions of other latent-state models.

The analysis of what happens in cross-cultural negotiation is more open. Because of infrequent contacts, it is unlikely that cross-cultural interactions are at an equilibrium. In a somewhat artificial case, when the agents are not aware

that the other parties come from different cultures, they may continue to apply the strategy learnt for their own culture. Since such strategies are unlikely to be in an equilibrium across cultures, the agents are open for exploitation.

Even when agents can identify that others belong to a different culture, their strategy may rely on patterns learnt in the same-culture negotiation. These are modified by (i) their previous contact with the other culture, and (ii) their prejudice. The question how these factors combine, as well as to what extent the mutual prejudices are in an equilibrium is open for future research.

Acknowledgements. Authors received support through the ARO grant number W911NF-08-1-0301.

References

1. Camerer, C.: Behavioral Game Theory. Experiments in Strategic Interaction. Princeton University Press (2003)
2. Gelfand, M.J., Brett, J.M., eds.: The Handbook of negotiation: Theoretical andvances and cultural perspectives. Stanford University Press, Palo Alto, CA (2004)
3. Ficici, S., Pfeffer, A.: Modeling how humans reason about others with partial information. In: AAMAS-08. (2008)
4. Ray, D., King-Casas, B., Montague, P.R., Dayan, P.: Bayesian model of behaviour in economic games. In: Advances in Neural Information Processing Systems 21. (2009)
5. Kuhn, H.W.: Extensive games and the problem of information. In Kuhn, H.W., Tucker, A.W., eds.: Contributions to the Theory of Games II. Annals of Mathematics Studies 28. Princeton University Press, Princeton (1953) 193–216
6. Koller, D., Milch, B.: Multi-agent influence diagrams for representing and solving games. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI). (2001) 1027–1034
7. Forges, F., von Stengel, B.: Computationally efficient coordination in game trees. Technical Report LSE-CDAM-2002-02, London School of Economics and Political Science, Centre for Discrete and Applicable Mathematics (2002)
8. Gordon, G.J., Greenwald, A., Marks, C.: No-regret learning in convex games. In: Proceedings of the 25th International Conference on Machine Learning. (2008)
9. Dudík, M., Gordon, G.J.: A sampling-based approach to computing equilibria in succinct extensive-form games. Preprint, submitted for peer review (2009)

Are All Trust Violations the Same? A Dynamical Examination of Culture, Trust Dissolution, and Trust Recovery

C. Ashley Fulmer
Michele J. Gelfand, Ph.D.
University of Maryland, College Park
1147 Biology/Psychology Building
College Park, MD 20742

Abstract. As our global interdependence grows, understanding how culture affects trust and how we can manage trust in intercultural relations is imperative. However, relatively few studies have focused on the relationship between trust and culture, and little of this work examined multiple trust phases sequentially to reveal the dynamics of trust over time. This research examined how the cultural differences of self-construal (individualistic vs. collectivistic) moderates the relationship between trust violation magnitude and trust change in two post-violation phases: trust dissolution and trust recovery. We adopted an economic game methodology, the Investment Game, which allows repeated measures to examine trust dynamics. The results revealed a joint effect of self-construal and trust violation magnitude on the dynamic of trust changes. Implications for intercultural negotiation will be discussed.

Keywords: Trust (Social Behavior), Cross-Cultural Differences, Dynamical Process.

1 Introduction

Trust has long been a focal interest in social sciences and linked to a myriad of social-psychological phenomena. It has been shown to facilitate interpersonal relationships [1], cooperation [2], teamwork [3], and leadership [4]. Furthermore, scholars view trust as a driving force in conflict de-escalation [5], a foundation for democracy [6], and a key driver of national economic well-being [7].

Moreover, as our global interdependence grows, interpersonal and institutional relations frequently cross national and cultural boundaries. However, there are surprisingly few empirical studies on trust in relation to cultures ([8]; for notable exceptions, see [9], [10]). In fact, it is estimated that over 90% of psychological research is conducted on less than 30% of the world population [11]. Understanding how culture affects trust is critical. It is unlikely that the trust process is universal, especially when many fundamental psychological phenomena have been shown to exhibit cultural specificity [12]. The first goal of this research project, therefore, was to examine how culture, in combination with situational factors, affects the trust process dynamically.

In addition to a lack of understanding on culture and trust, the trust literature is also limited by its predominant focus on trust building [8]. There has been a growing concern about the prevalence of trust violations [13], [14] but little is known about their impact on relationships. The second goal of our study was to focus on the two trust phases after violations, namely trust dissolution and trust recovery. *Trust dissolution* refers to a period when, after trust violations, trustors decide to lower their trust in trustees, and *trust recovery* occurs when trust stops declining after violations and starts to rebound [15].

Despite the recognition of multiple trust phases in the literature, existing research tends to be phase-specific, examining one trust phase at a time [15]. We argue that such a narrow focus provides only a snapshot of the trust process as it naturally occurs. Interactions between individuals are continuous, and the isolation of a single trust phase cannot provide a holistic picture of how trust patterns unfold over time. For example, after a violation, trust may plummet initially but rebound with repeated interactions [15]. Thus, the third goal of our study was to examine trust dissolution and trust recovery sequentially.

In sum, to begin to fill the theoretical and empirical void in trust literature, our research 1) examines the impact of culture on trust in conjunction with a situational variable, 2) focuses on the trust after violations, and 3) measures trust continuously in trust dissolution and recovery to reveal its dynamics after trust violations. Further, we believe that the fluctuation in trust across phases is influenced by situational factors, cultural differences, and interactions between these factors. For example, the magnitude of trust violations is expected to affect trust dynamics, with large violations leading to faster trust dissolution and slower trust recovery than small trust violations. However, these trust patterns are also expected to differ across cultures and, as we discuss below, cultural factors are expected to interact with the magnitude of violations. To achieve these three goals of our research project, we adopted the paradigm of the Investment Game (IG) [16]. The IG collected repeated measures that allow longitudinal modeling to reveal the nonlinear and dynamic nature of trust.

1.1 Trust Violations

The inclusion of trust violation in examination of trust dynamics is imperative, as the act of trusting implicitly accepts future uncertainty and risk [9], [15], [10]. In fact, the very conditions that foster trust, and the existence of trust itself, allow for malfeasance [17].

Trust violations vary in their degree; a delay in returning a book is clearly different from failure to keep a marriage vow. Small transgressions, therefore, should not have the same impact on trust changes as large breaches of trust. For example, Tomlinson, Dineen, and Lewicki [18] found the magnitude of a violation moderated the relationship between the estimated likelihood of future violations and trust recovery. While it is reasonable to expect that large trust violations will lead to faster trust decline and slower trust recovery than small violations, a key question is how cultural influence affects this trust pattern. Thus, our research project examined the moderating effect of culture on the relationship between violation magnitude and trust dynamics.

1.2 The Relationship between Trust Violation and Self-Construal

In this research project, we examined the interplay between trust and culture through the construct of self-construal. Individuals have divergent views about the self and others [19]. Markus and Kitayama [12] proposed that people with individualistic self-construal endorse "a conception of the self as an autonomous, independent person" (p. 226). This conceptualization of the self is dominant in the West. In contrast, people in Asian and Middle Eastern cultures tend to have collectivistic self-construal and view "the self. . . not as separate from the social context but as more connected and less differentiated from others" (p. 227).

Therefore, trustors with collectivistic self-construal should, generally speaking, identify with their trustees more than do trustors with individualistic self-construal. Social identity theory asserts that when individuals identify with another, they are motivated to maintain positive perceptions of the person to maintain high self-esteem [20]. This motivation, combined with the higher level of trustworthiness individuals perceive from people with whom they identify [21], should prompt collectivistic trustors to be more tolerant toward minor trust violations and restore trust more easily than individualistic trustors.

> *Hypothesis 1: After **small** trust violations, collectivists will experience **slower** trust decline and **faster** trust recovery than individualists.*

With large trust violations, however, we propose that the pattern would be reversed. Because of their deeper identification with their social context, it may be more difficult for collectivists to overcome large trust violations than individualists. Evidence of the "black sheep effect" [22] has demonstrated that, when identified others exhibit major shortcomings, people can have low tolerance of the failings and engage in denigration of these individuals. This black sheep effect is the strongest when individuals *closely identify with others* [23], such as a trustor with collectivist self-construal. The combination of large violations and collectivistic self-construal, therefore, should lead trust breaches to be even more personally relevant to collectivistic trustors. Consequently, collectivistic trustors would attempt to distance self from the betrayer more by considerably decreasing their trust to another and take longer to recover from the damage of large violations than individualistic trustors.

> *Hypothesis 2: After **large** trust violations, collectivists will experience **faster** trust decline and **slower** trust recovery than individualists.*

2 Methods

2.1 Design and Participants

The study employed a 3 x 2 design, examining the processes in which trust violations (large vs. small vs. control) and self-construal (collectivistic vs. individualistic) affect

trust dissolution and trust recovery among student samples [24]. A total of 69 students in a large, public university participated in the study. Of these, 29% were male and 71% were female. Additionally, 15% were White, 10% were Asian American, 55% were African American, and 20% were Hispanic. The mean age was 19.67 (SD=1.35).

2.2. Apparatus and Procedure

We conducted laboratory experiments using a variant of the Trust Game, the Investment Game (IG) [16], to measure trust. The IG is ideal for our study for a number of reasons. First, the design of IG affords social exchanges that mirror real-world interactions. Second, the structure of IG allows examination of how trust violations affect trust changes. Finally, an iterated IG [25] is suitable in examining nonlinear attitudinal changes because it permits repeated measures of trust.

In each experimental session, participants played 19 rounds of IG on computers in individual rooms. The number of rounds, determined in the pilot studies, balances the time per round and the total number of rounds necessary to observe the nonlinear and dynamic trust changes. Upon starting the experiment, participants were informed that they will engage in multiple rounds of brief interaction with another participant. In actuality, participants played the IG with the computer-programmed partner. Appendix A described the game structure and specific steps involved in details. Participants filled out a self-construal scale after completing the IG.

2.2. Measures

Two measures of trust, behavioral and attitudinal, were collected at each round during the game. The behavioral measure of trust was represented by the number of coins participants allocated to the partner and the attitudinal measure was assessed through a single item "how much do you trust the other player?" on a 7-point scale (1 = *not at all*, 7 = *completely*). Self-construal was measured using the scale of independent and interdependent self-construal [26]. On a 7-point scale (1 = *strongly disagree*, 7 = *strongly agree*), 24 items measured the extent to which participants have independent versus interdependent self-construal. A sample item was "I enjoy being unique and different from others in many respects." The alpha for this scale was .71.

3 Results

All data analyses were conducted using the R software environment for statistical computing [27] and the nlme package [28]. Specifically, we applied growth modeling [29] to analyze the data, a common method in longitudinal data analysis. Furthermore, we conducted a mixed effect model, setting rounds, violations magnitude, and collectivistic self-construal as fixed effects while allowing for within-individual random variation in slopes associated with rounds of the IG. The model included 2 levels. The higher level consisted of individuals' collectivistic self-construal and their

attitudinal and behavioral measures of trust, while the lower level consisted of rounds. Our random coefficient model (RCM) following the recommended procedures [30].

Our results showed that a significant three-way interaction effect of time (as rounds), violation magnitude, and levels of collectivistic self-construal affected the amount of coins participants allotted to their programmed partner (b=-0.68, t(1238)=-1.97, p<.05). Supporting Hypothesis 1, collectivistic trustors displayed slower trust dissolution and faster trust recovery after small trust violations than individualistic trustors.

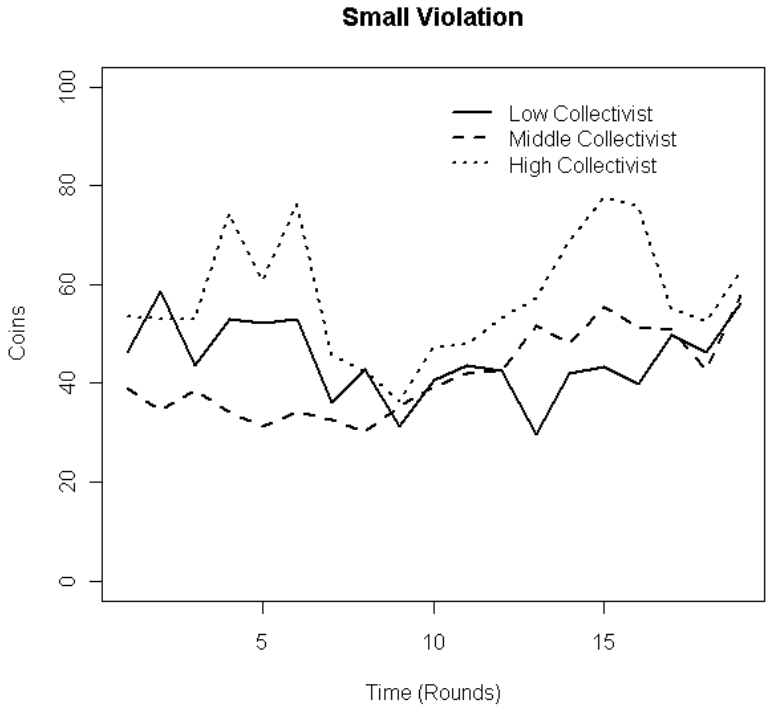

Fig. 1. Trust patterns of individuals with high, medium, and low levels of collectivistic self-construal after small trust violations.

In contrast, collectivistic trustors displayed faster trust dissolution and slower trust recovery after large trust violations than individualistic trustors. This pattern supported Hypothesis 2.

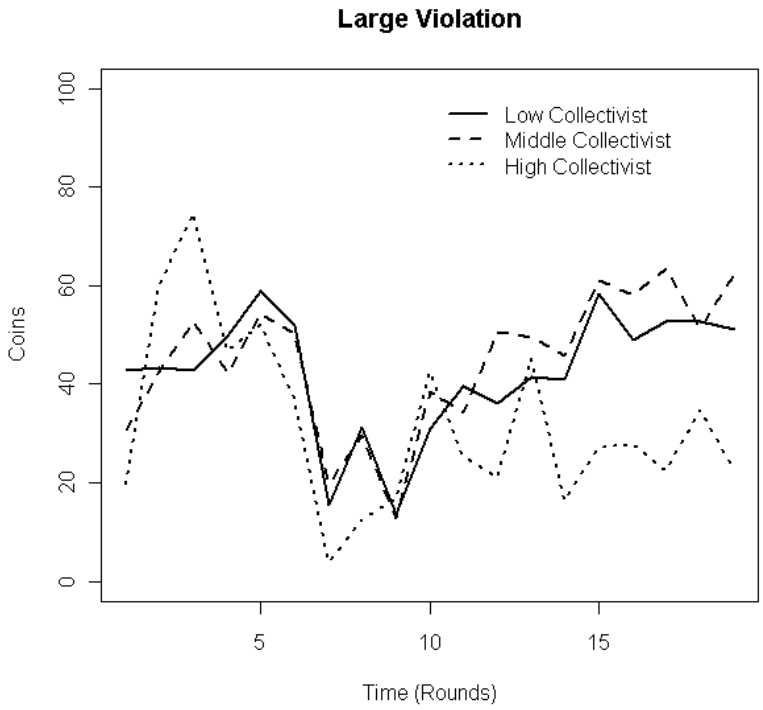

Fig. 2. Trust patterns of individuals with high, medium, and low levels of collectivistic self-construal after large trust violations.

4 Discussion

Prior research has conceptualized the trust process as consisting of three distinct phases [15]. In the present study, we focused on trust dissolution and trust recovery. In addition, we investigated whether the cultural differences of collectivistic and individualistic self-construal interacted with the situational factor of trust violation magnitude on these two phases. Results from the discontinuous growth modeling revealed a significant three-way effect among rounds, self-construal, and magnitude of trust violation. Further, supporting our hypotheses, we found that collectivistic trustors display divergent patterns of trust dissolution and trust recovery depending on the magnitude of trust violations. After small trust violations, collectivists showed

slower trust decline and faster trust recovery than individualists. In contrast, after large trust violations, collectivists showed faster trust decline and slower trust recovery than individualists. In other words, compared to individualistic trustors, collectivistic trustors tend to allow more latitude for small trust violations but were less flexible with large trust violations. The results with large violation indicated an existence of the black sheep effect among collectivistic trustors, exhibiting low tolerance toward large trust violations and engaging in more negative behaviors toward the ingroup members who committed the large trust violations.

The contributions of our study are three-fold. First, in response to concerns about little research on trust violations [8], the present research focused on the trust process after violations. Second, our study was among the first to examine multiple trust phases sequentially. The findings support the notion that trust is dynamic, as suggested by other scholars in the field [15]. Finally, our study included both the cultural variable of self-construal and the situational variable of trust violation magnitude as antecedents of trust changes. The results on joint effects of the cultural and situational variables suggest that researchers need to take both factors into account to fully explore the complexity of trusting relationships. Future research should examine additional situational factors such as time pressure, accountability, and the nature of the group (e.g., friend versus stranger) along with cultural factors to predict the dynamical nature of trust.

Both trust and culture are increasingly important in determining our societal well-being. We believe that a dynamical approach to trust processes, combined with rigorous and appropriate methodology, will provide scientists better understanding of this important construct and make an impact on our field and our global community.

Acknowledgments

This research is based upon work supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-08-1-0144.

References

1. Rempel, J.K., Holmes, J.G., Zanna, M.P.: Trust in Close Relationships. Journal of Personality and Social Psychology, 49, 95--112 (1985)
2. Buchan, N.R., Croson, R.T.A., Dawes, R.M.: Swift Neighbors and Persistent Strangers: a Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange. American Journal of Sociology, 108(1), 168--206 (2002)
3. Lawler, E.J.: Affective Attachments to Nested Groups: a Choice-Process Theory. American Sociological Review, 57(3), 327--339 (1992)
4. Dirks, K.T., Ferrin, D.L.: Trust in Leadership: Meta-Analytic Findings and Implications for Research and Practice. Journal of Applied Psychology, 87, 611--628 (2002)
5. Axelrod, R.: The Evolution of Cooperation. Basic Books, New York (1984)

6. Putnam, R.D.: Making Democracy Work: Civil Traditions in Modern Italy. Princeton University Press, New Jersey (1993)
7. Fukuyama, F.: Trust: The Social Virtues and the Creation of Prosperity. Free Press, New York (1995)
8. Schoorman, F.D., Mayer, R.C., Davis, J.H.: An Integrative Model of Organizational Trust: Past, Present, and Future. Academy of Management Review, 32(2), 344--354 (2007)
9. Bohnet, I, Herrman, B., Zeckhauser, R.: The Requirements for Trust in Gulf and Western Countries. Quarterly Journal of Economics *(in press)*
10. Yamagishi, T., Cook, K.S., Watabe, M.: Uncertainty, Trust, and Commitment Formation in the United States and Japan. American Journal of Sociology, 104, 165--194 (1998)
11. Triandis, H.C.: Cross-Cultural Industrial and Organizational Psychology. In: Handbook of Industrial and Organizational Psychology, pp. 103--172. Consulting Psychologists Press, California (1994)
12. Markus, H.R., Kitayama, S.: Culture and the Self: Implications for Cognition, Emotion, and Motivation. Psychological Review, 98, 224--253 (1991)
13. Elangovan, A.R., Shapiro, D.L.: Betrayal of Trust in Organizations. Academy of Management Review, 23(3), 547--566 (1998)
14. Morris, J.H., Moberg, D.J.: Work Organizations as Contexts for Trust and Betrayal. In: Citizen Espionage: Studies in Trust and Betrayal, pp. 163--187. Praeger Publishers/Greenwood Publishing Group, Connecticut (1994)
15. Rousseau, D., Sitkin, S., Burt, R., Camerer, C.: Not So Different After All: a Cross-Discipline View of Trust. Academy of Management Review, 23, 393--404 (1998)
16. Berg, J., Dickhaut, J., McCabe, K.: Trust, Reciprocity, and Social History. Games and Economic Behavior, 10, 122--142 (1995)
17. Granovetter, M.: Economic action and social structure: The Problem of Embeddedness. American Journal of Sociology, 91, 481--510 (1985)
18. Tomlinson, E.C., Dineen, B.R., Lewicki, R. J.: The Road to Reconciliation: Antecedents of Victim Willingness to Reconcile Following a Broken Promise. Journal of Management, 30(2), 165--187 (2004)
19. Triandis, H.C.: The Self and Social Behavior in Differing Cultural Contexts. Psychological Review, 96, 506--520 (1989)
20. Tajfel, H., Turner, J.C.: An Integrative Theory of Intergroup Conflict. In: Austin, W. G., Worchel, S. (Eds), The Social Psychology of Intergroup Relations. Brooks/Cole, California (1979)
21. Brewer, M.B., Kramer, R.M.: The psychology of Intergroup Attitudes and Behavior. Annual Review of Psychology, 36, 219-243 (1985)
22. Marques, J.M., Yzerbyt, V.Y., Leyens, J.-P. : The Black Sheep Effect: Judgmental Extremity toward In-Group Members as a Function of Group Identification. European Journal of Social Psychology, 18, 1--16 (1988)
23. Branscombe, N., Wann, D., Noel, J., Coleman, J.: In-Group or Out-Group Extremity: Importance of the Threatened Social Identity. Personality and Social Psychology Bulletin, 19(4), 381--388 (1993)
24. James, W.L., Sonner, B.S.: Just Say No to Traditional Student Samples. Journal of Advertising Research, 41(5), 63--71 (2001)

25. Cochard, F., Nguygen Van, P., Willinger, M.: Trusting Behavior in a Repeated Investment Game. Journal of Economic Behavior & Organization, 55(1), 31--44 (2004)
26. Singelis, T.M.: The Measurement of Independent and Interdependent Self-Construals. Personality and Social Psychology Bulletin, 20, 580--591 (1994)
27. R Core Team: R: A Language and Environment for Statistical Computing (Version 2.8.1) [Computer software]. Vienna: R Foundation for Statistical Computing (2004)
28. Pinheiro, J.C., Bates, D.M., DebRoy, S., Sarkar, D., R Core Team.: nlme: Linear and Nonlinear Mixed Effects Models (Version 3.1–89) [Computer software]. Vienna: R Foundation for Statistical Computing (2008)
29. Singer, J.D., Willett, J.B.: Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press, New York (2003)
30. Bliese, P.D., Ployhart, R.E.: Growth Modeling Using Random Coefficient Models: Model Building, Testing, and Illustrations. Organizational Research Methods, 5, 362-–387 (2002)
31. Lount, R.B., Zhong, C., Sivanathan, N., Murnighan, J.K.: Getting Off on the Wrong Foot: Restoring Trust and the Timing of Breach. Personality and Social Psychology Bulletin, 34(12), 1601--1612 (2008)
32. Sitkin, S.B., Roth, N.L.: Explaining the Limited Effectiveness of Legalistic 'Remedies' for Trust/Distrust. Organization Science, 4(3), 367--392 (1993)
33. Murnighan, J.K.: Defectors, Vulnerability, and Relative Power: Some Causes and Effects of Leaving a Stable Coalition. Human Relations, 34, 589--609 (1981)

Appendix: Structure of the Investment Game (IG)

The IG involves two players, Player A and Player B. In this study, all participants were assigned to the role of Player A (trustor) and the computer-programmed partner was Player B (trustee). In the beginning of each round, Player A was given 100 coins and decided a proportion of the endowed coins to entrust to Player B (0-100). This decision revealed how much Player A trusts Player B. After allocating coins to Player B, Player A also indicated their level of trust in Player B. Thus, in addition to a behavioral measure of trust as represented by the entrusted coins, the study included an attitudinal measure of trust with the question "how much do you trust the other player?" (7 = *completely*; 1 =*not at all*).

The amount sent to Player B by Player A was then tripled, and Player B decided a portion of the tripled coins to return to Player A. The game constituted of 19 such round. In violation conditions, trust breaches occurred in the 6^{th}, 7^{th}, and 8^{th} rounds trust violations. Violations occurred during these rounds because they were relatively early in the game, and yet were not right in the beginning, which may lead to irreversible damage to trust [31]. The three rounds of violations were designed so that participants would not perceive the violations as an isolated incident, which the participant could discount and keep trust unaffected [32]. During the three violation rounds, Player B kept either all or the majority of the coins, depending on the violation conditions. In non-violation rounds, Player B returned approximately half of

the tripled coins with small random variations. At the end of the 19th round, the game stopped. As knowledge of the end of a transaction tends to decrease cooperation [33], participants did not know how many rounds remained during the game before the end.

POMDP based Negotiation Modeling

Praveen Paruchuri, Nilanjan Chakraborty, Roie Zivan, Katia Sycara, Miroslav Dudik, Geoff Gordon
{paruchur, nilanjan, zivanr, katia, mdudik, ggordon}@cs.cmu.edu

School of Computer Science, Carnegie Mellon University

Abstract. As the world gets increasingly connected, business and political negotiations need to happen not just between people of similar cultural background but also across people of different cultures. An agent-based computational model of negotiation would help in understanding and improving the inter-cultural negotiation process. A major challenge in the development of such autonomous agents lies in developing the reasoning model of the agent. In this paper, we discuss the issues and challenges of developing a POMDP-based agent model for inter-cultural negotiation. POMDPs are promising for the following reasons: (a) POMDPs provide a decentralized way of solving the problem which is an inherent characteristic of the negotiation domain. (b) POMDPs provide a natural way to capture the sequential nature of the bargaining process i.e. they capture the process rather than just focusing on the outcome. (c) POMDPs can express the various important factors that affect negotiation such as culture.

1 Introduction

Negotiation is an interactive process by which multiple parties with limited common knowledge try to arrive at an agreement over a set of issues with possibly conflicting preferences over the issues. The topic of negotiation has received wide attention across various fields such as political science [8], economics [17], behavioral sciences [21], computational sciences [14] etc. For example, political scientists are interested in analyzing and predicting the negotiation processes between various countries and cultures while computer scientists are more interested in building computational models that can be embodied in automated agents. Most work in computational modeling to-date has focused on the outcome of a negotiation. In this paper, we propose a computational modeling approach for a general purpose negotiation problem with special emphasis on capturing the process of negotiation. While the approach presented is general, we pay special attention to the cultural aspects of the problem and show that our approach has a natural way of capturing such factors. Numerous studies (see [10] and references therein) have established that culture plays a crucial role in the way a negotiation progresses and our aim is to be able to evaluate these effects in a formal framework.

Game-theoretic techniques are an important class of computational techniques that have been used to study negotiation (see Section 5 for a discussion of other techniques). Most game-theoretic models in strategic interactions among self-interested agents (negotiation is an instance of this) aim to find solutions in terms of equilibrium point concepts

(e.g., Nash equilibria). However, it may be computationally intractable to find such an equilibrium point. Moreover, it has been noted that, in a negotiation scenario, people may not follow the strategy corresponding to the game-theoretic equilibria [7]. Thus, traditional game-theoretic techniques are not suitable for modeling the evolution of the negotiation process. To capture the evolution of the negotiation process, we model the negotiating agents as a dynamical system that evolves in time. In particular, we model the negotiation problem in a decision-theoretic framework as a Partially Observable Markov Decision Process (POMDP) [12]. The advantages of the POMDP-based modeling approach are as follows: (a) Our approach is decentralized, i.e., each agent solves her own POMDP model while maintaining a belief about the other agents. This is in contrast to most game theoretic approaches where the payoffs of all the agents are assumed to be common knowledge. (b) In contrast to most computational models that are concerned only with the outcome, POMDPs provide a natural way to capture the sequential nature of the process while incorporating the new observed data such as the other agent's actions; additionally POMDPs provide ways to refine an agent's belief about other agents. (c) POMDPs can incorporate the effect of cultural factors in a natural way, e.g., given the action of the opponents and their cultural background, it uses the knowledge in interpreting the actions of the opponent, or in deciding the best action to take (please see the discussion in the next paragraph). The purpose of this paper is to illustrate with a simple negotiation example the construction of the POMDP model for negotiation. We also point out the various challenges in the development of a POMDP model (see Section 3).

Figure 1 illustrates a common way of representing how cognitive schemata [26] change during negotiations. The figure shows how schema's filter and interpret incoming stimuli and guide outgoing reactions for a simple two-party (say, A and B) interaction, e.g., for resolving a conflict. Party A's schema enters in two places. First, it is the lens through which party B's behavior will be interpreted and second it is the filter through which A's actual intentions will give rise to concrete behaviors visible to the other party. In this way, schema's become relevant whenever an individual is taking information from the outside world or offering behavior to the outside world. A's culture and history of interaction with B (or members of B's culture) will influence party A's schema. Important components of a schema are goals (what is appropriate to try to achieve), norms (what is appropriate behavior to go about getting what you want), and beliefs and attributions about the character of the other person . A's schema includes "who B is", which influences A's interpretation of B's behavior "what B is doing". This drives A's intentions or strategies for subsequent moves (e.g., should A be cooperative or not). A's intentions will then drive A's behavior, as filtered again through A's schema, which, includes norms for appropriate behavior. Since this is a symmetric situation, B's schema filters A's behavior and influences how B interprets that behavior, which influences how B intends to respond.

We now provide a brief intuition on how the cognitive schema presented in Figure 1 naturally fits into a POMDP model. To illustrate this, we provide here an informal description of the POMDP model while a more formal description is provided in Section 3. Figure 2 (taken from [12]) shows the working of a POMDP. The world of the POMDP

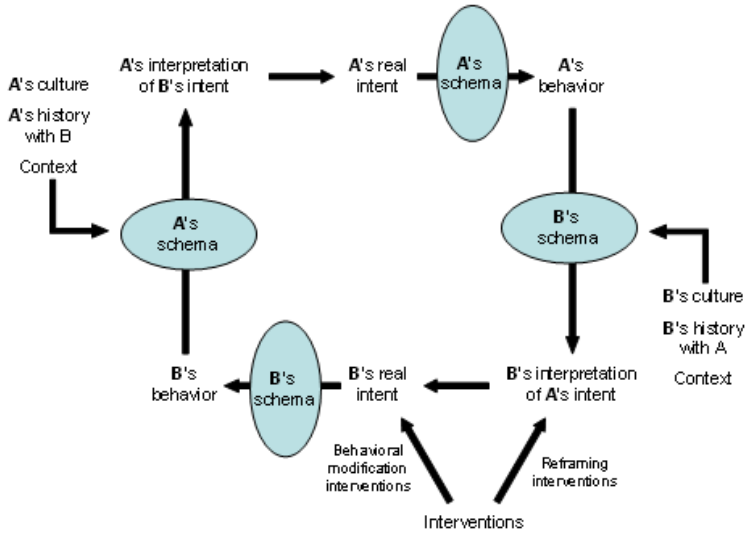

Fig. 1. Cognitive schema in dynamic collaboration and negotiation.

is composed of states. Initially, an agent believes that it is in a particular state or has a probability distribution over states, called the belief state. The agent takes an action and gets an observation of the new state she reaches. Given that the new state may not be directly inferred from the observation, the state estimator (labeled SE) derives the new belief state based on the last action, the current observation and the previous belief state. Once the new belief state is calculated, the agent takes a new action and the process continues till an end state is reached. The block labeled π represents the POMDP policy. The policy of the POMDP maps a belief state to an action. Informally, the policy is a table which can be computed beforehand that maps a belief state (and hence an observation) to the optimal action. We now model the POMDP from the perspective of party A for the cognitive schema presented in Figure 1. The context consisting of A's culture and A's history with B directly maps to the initial belief of a POMDP model, i.e., the context specifies the probability distribution over the states at the start. A's interpretation of B's intent and the various possible interventions map to the observation in a POMDP, while A's schema and real intent becomes part of the state space. A's behavior gets captured in the action set of the POMDP. The perceived model of B is also represented in the state space of A's POMDP. Similar mapping can be done for agent B. This direct mapping between a general purpose cognitive schema for negotiation and a POMDP model, reinforces our belief that modeling the negotiation problem as a POMDP may be a good approach to follow.

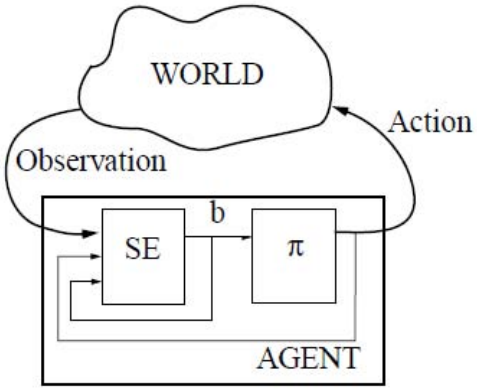

Fig. 2. Working of a POMDP

2 A general purpose negotiation setup

We now introduce the general purpose negotiation setup that we consider in this paper. In Section 3.1, we will present a simple example to illustrate the POMDP based model discussed below. Consider $i \in \{1, 2, .., n\}$ self-interested agents negotiating a set of issues $j \in \{1, 2, .., m\}$. The set of issues being negotiated can be of different kinds, some involving simple price bargaining while others involve more complex dialogue based negotiations. Given that each agent is inherently different, they have different expectations on what a fair solution is, what a fair way to negotiate would be etc. The differences between each agent can be captured through the notion of types. We assume that each agent has a type $T \in \{t_1, t_2, .., t_n\}$ based on which she acts. For purposes of this paper, we assume that an agent knows its type T while it has a probability distribution over the opponent types $T' \in \{t'_1, t'_2, .., t'_n\}$. The type of an agent is derived as a function of numerous individual factors. One such factor can be the agent's personality which can be classified as either *Selfish* or *Altruist*. While this is a coarse representation, a finer representation can involve mixtures in various proportions of these two personalities. Therefore, if a personality value of 0 represents Selfish agent and 1 represents an Altruist, 0.5 will represent an *Equality* opponent, namely a player who is interested in obtaining value both for herself and the other player. Other possible factors that can be included in the agent's type include the agent's motivation and the history of negotiations of the agent. The type space of an agent can then be obtained as a cartesian product over the set of all the factors.

We also assume that the agents know the cultures of all the other agents at the start of the game. As described in [5], various cultural factors such as individualism and collectivism, egalitarianism versus hierarchy, direct versus indirect communication and other factors play a significant role in forming the initial model of the opponent especially in the absence of any other significant individual-tailored information. Therefore, culture may be modeled as skewing the prior probability over the opponent types and

hence helps in building a realistic initial belief that an agent has about the other party. It can happen that over the course of negotiation an agent realizes that the initial belief it held about the other agent may not be true and hence refines the model as the negotiation progresses. We now describe how the POMDP framework captures the general negotiation process just presented.

3 The Partially Observable Markov Decision Process Framework

In Section 1 we provided an informal description of the POMDP model. Formally, a POMDP can be represented as the tuple $\{S, A, T, \Omega, O, R\}$, where S is a finite set of states; A is a finite set of actions; $T(s, a, s')$ provides the probability of transitioning from state s to s' when taking action a; Ω is a finite set of observations; $O(s', a, o)$ is the probability of observing o after taking an action a and reaching state s' and $R(s, a)$ represents the reward function. A belief state b, is a probability distribution over the set of states S. A value function over a belief state is defined as: $V(b) = max_{a \in A} R(b, a) + \beta \sum_{b' in B} T(b, a, b') V(b')$. Once the negotiation problem is cast into the POMDP framework, many algorithms both heuristic and exact exist in the literature to find the approximate/optimal POMDP policies [12], [20]. A policy here refers to the mapping between a belief state (and hence the observation at that time-step) to an action. Effectively, the agent solves the POMDP and obtains a policy table. The agent can then use this table to decide on the appropriate action to take at each time-step based on the observation the agent obtains at that time-step and the state the agent is in.

The main challenge in casting a negotiation as a POMDP lies in defining the tuple $\{S, A, T, \Omega, O, R\}$ for an instance of the problem. The state space S can be defined using the knowledge of the problem domain and the various factors affecting negotiation that has been identified in the behavioral sciences literature [21]. The action space A and the space of observations Ω can also be formed using domain knowledge as well as knowledge about the strategies used by people in a negotiation (that has been identified in the behavioral sciences domain). From the point of view of modeling, the main hurdle lies in coming up with the appropriate parameters for the state transition function, T, observation function O, and the reward function R. One way to overcome this is to conduct negotiation experiments using test-subjects and design the experiments in such a way so as to extract the parameters using machine learning techniques.

The solution of the POMDP model above will be a policy that prescribes the action an agent should take given the state of the world she believes that she is in. In general, finding an optimal policy, i.e., a policy that optimizes the expected value of R may not be computationally tractable. However, for purposes of this work, we may not need to solve the complete POMDP. Instead we assume that our agent has an initial belief of the model of the other negotiating agents based on cultural and other appropriate factors. If such an assumption can be made, which is true in our domain since the culture of the negotiating parties is assumed to be known, the POMDP becomes easier to solve,

by converting to a belief MDP. Here, the belief MDP refers to a MDP where the set of states were derived by calculating all the possible belief states of the POMDP.

Another generic concern about modeling using a POMDP is that the POMDP policy is deterministic, i.e., for a given belief state, the action taken by the agent is unique. Thus, it is theoretically possible for an opponent to learn the negotiating agent's policy and exploit it. However, in the negotiation domain, it is an open problem whether it is practically possible for the opponent to exploit a POMDP playing agent. Moreover, efficient algorithms that do a controlled policy randomization for countering such deception tactics have been developed for belief MDPs and we plan to utilize them in our work (see [19] for details).

We now summarize the steps for finding the optimal POMDP policy before getting into the modeling details. Informally, given an initial belief, solving a POMDP involves the following steps:

- Convert the POMDP to a belief MDP by enumerating all the possible belief states and applying the other relevant transformations [12].
- The newly obtained (belief) MDP can be solved efficiently using the standard Value or Policy iteration algorithms [3] to obtain the policy table.
- The agent can then use this policy table to map the newly obtained observation to a relevant action in real-time.

3.1 Simplified negotiation problem in a POMDP framework

We now present a simple negotiation example to illustrate various aspects of the POMDP model. We consider a transactional negotiation scenario where two agents are negotiating over the price of a single item. Agent 1 is assumed to be the seller and agent 2 the buyer. We assume that the best price for the buyer is 0 while it is 10 for the seller. Note that this scale captures a general set of scenarios since any other scale can be normalized and shifted to fit in this. The corresponding worst case scenario is 10 for buyer and 0 for the seller. We assume that a single factor namely the personality determines the type of all the agents. As defined earlier we model personality via a value between 0 to 1, where 0 corresponds to Selfish type and 1 corresponds to Altruist. Any other intermediate value represents a mixture between these two personality types. We assume that both the agents know each other's culture. In the following subsection, we present the seller's POMDP for this simple example.

3.2 World States: S

S is the set of world states. For the negotiation problem, we model the state of the POMDP as the following vector: $< MyType, OpponentType, CurrentNegotiationState >$. In our simplified example, $MyType$ (corresponding to seller) is a single number ranging between 0 to 1 based on my personality. For explanation purposes we focus on a

particular state in which the agent is of Equality type. Assigning a value to $MyType$ becomes harder as additional factors get added such as motivation, history of negotiation etc. The second component in our POMDP model is the $OpponentType$ (corresponding to buyer) which is again modeled as a single number since we consider a single factor. Note that there can be infinite values that this single number can take since personality is a continuous variable. For simplicity of state representation, we discretized the personality factor to lie in the set (of 11 values) $\{0, 0.1, 0.2, .., 0.9, 1\}$. If a new factor gets added, the set of new types can be obtained as a cartesian product of the old types and the set of values for the new factor. The third component of our state is the $CurrentNegotiationState$. We represent this using the following vector: $< MyPrice, OpponentPrice >$. $MyPrice$ is the current price that I want to sell the item for while the $OpponentPrice$ refers to the price the buyer wants the item for. For simplicity of representation, we discretized both $MyPrice$ and $OpponentPrice$ to lie in the set $\{0, 1, .., 9, 10\}$. The price scale can be enlarged or shrunk based on the domain. The set of states can then be obtained as a cartesian product of the set of values of each component of the state vector. Therefore, the total number of possible states for this representation is $1 * 11 * < 11 * 11 >$ = 1331.

3.3 Action set: A

The simplest way to construct the action set A for our simplified problem is to model the fact that each agent can negotiate for any of the 11 possible values at any time instance. However, this makes the POMDP harder to solve. We therefore make the following simplifying assumptions without necessarily bringing in any restrictions. In particular, we assume that if an agent quotes a price x, she will either remain at x or increase/decrease the current price by an integer z. Setting z to 2, we obtain that an agent can either remain at the same price or increase or decrease the price by a maximum of 2 units. The new action set corresponding to this would then be the set $\{Same, Concede1, Concede2, Retaliate1, Retaliate2, End\}$. Here, $Same$ would mean that the agent remains at the same price, $Concede1$ would mean add 1 to current price for buyer or subtract 1 for the seller while $Retaliate1$ would mean subtract 1 from the current price for the buyer while add 1 for the seller. End would mean that the agent agrees to the current price and the deal is closed. The action set gets complicated once we consider the fact that agents can have dialogues instead of a set of numbers.

3.4 Transition Function: T

A transition represents the probability with which an action a taken from a state s leads to a state s'. Transitions in our domain are stochastic. This is because when an agent takes an action say $Concede1$, it cannot be sure what the opponent's action would be and hence the state it reaches (which includes the opponent's current negotiation price). Figure 3 provides a pictorial description of our transition function. The leftmost state represents the state under consideration. In this state, $Mytype$ is 0.5 and the ground truth of the opponent's type is also 0.5. The current negotiation price for myself (the

seller) is 10 while that of the buyer is 0. If the seller takes an action $Concede1$, i.e., seller proposes a new price of 9, the buyer can respond by remaining at 0 or moving to 1 or 2. Given the ground truth that the buyer is of $Equality$ type, our transition function would capture this fact by assigning a low transition probability (0.1 here) to state that has the buyer's current negotiation price remaining at 0 while assigning a higher probability (0.4 here) to states that have the buyer's current negotiation price as either 1 or 2. Note that we also allow for a small transition probability (0.1 here) where the ground truth about the buyer can change, i.e., she can no longer be $Equality$ type. This is to account for the fact that the buyer's behavior need not be fixed and can be a function of time and the seller's price. While we do not consider in this paper, similar argument can hold for the seller and can be easily represented in the POMDP at the cost of increased state space.

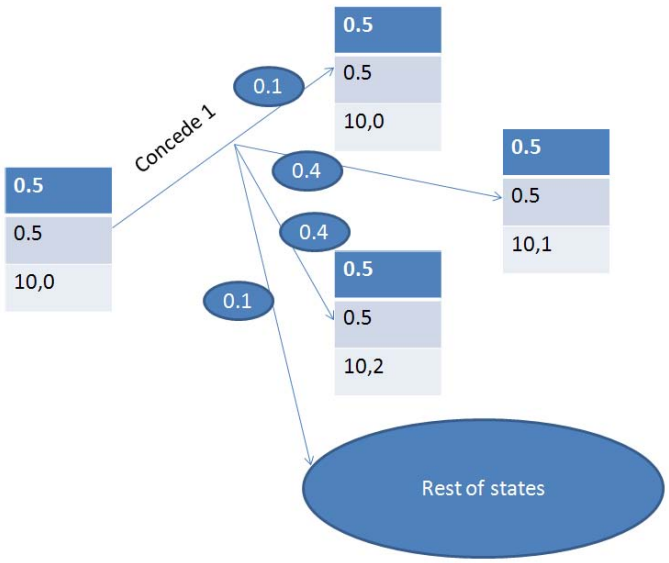

Fig. 3. A simple negotiation example encoded as a POMDP

3.5 Observations: Ω

The set of observations that an agent (seller here) observes are the actual prices quoted by the opponent. Therefore, there are 11 possible observations in the domain corresponding to the 11 possible prices the buyer can quote.

3.6 Observation function: O

Assuming that there is no noise in the observations i.e. if an agent quotes price p, the other agent actually sees that the quotation is p, the observation function is deterministic. This is because we model the quoted price as part of the state. Therefore, for each $P(o/s', a)$ = $P(ObservingPrice(x)/StateWithBuyerPrice(x), a)$ = 1 and 0 otherwise.

3.7 Reward function: R

There can be multiple ways to model the reward function. One potential way we consider here is: The reward is calculated as a function of the agent's type, its current negotiation price and the action it takes. For example, when a selfish agent at current price 0 takes action $Concede1$, she gets a small positive reward as opposed to a high negative reward incurred when the current price is 4.

4 Effect of culture on modeling the agents using POMDP

POMDP's are a standard framework to represent domains in which agents have partial observations of their surroundings. One example can be a robot which uses its sensors to determine its location while trying to reach a target [24]. In our negotiation domain, each agent has its own type. A negotiating agent can only observe the actions of its opponent but the true type of the opponent remains hidden. The POMDP framework allows representation of this partial knowledge and accounts for this uncertainty while calculating the optimal negotiation strategy. An important factor that determines the type of the opponent is her culture. The rest of this section focuses on how cultural factors affect the POMDP modeling.

4.1 Including culture in modeling initial beliefs

In the modeling of a negotiating agent as a POMDP, the agent knows information such as her type and the offers made by both parties but does not know exactly the type of the opponents. Therefore, the agent maintains a probability distribution over the opponent types as a possible model of the opponent. At the start of the negotiation, the agent would not have observed any opponent actions. In the absence of any information, one possible initial model of the opponent would then be to maintain a uniform distribution over all the types. This means that the agent has an initial belief which is a uniform distribution over all the states with negotiation value 0 for buyer and 10 for seller. However, this is not the case when we take culture into consideration. If the agent is negotiating with an agent from the same (her own) culture, we would expect the agent to model the opponent more accurately which strongly skews the initial belief. When the negotiation includes agents from different cultures, the agent's initial beliefs are usually a stereotype of the opponent's culture [26].

4.2 Effect of Culture on agents' observations

Culture has a dramatic effect on the observations made by the negotiating agent throughout the negotiation process. A negotiating agent, after taking an action, gets the opponents action as the observation. In our simplified example, the negotiating agent observes the price quoted by the opponent. The interpretation of this observation refines the agent's model of the opponent. For example, if the agent modeled her opponent as an altruist initially, it can later refine the opponent model as the negotiation progresses to account for observations that make the opponent seem selfish. While this refining process can correct initial bias, culture may play a significant role here. For example, many studies on culture in negotiation have reported that different cultures express the same intentions differently (referred to as interventions in the cognitive schema). Therefore, when an agent makes an offer or utters a dialogue that is her own mind is expressing altruism, this observation can be interpreted wrongly by the opponent and the opponent's belief can be updated incorrectly.

4.3 Culture in States, Transitions and Rewards

We now show how cultural factors could affect the other components of the POMDP tuple namely states, transitions and rewards. For example if a self interested agent negotiates with an agent from a culture in which there is a high probability of altruism, there would be a low transition probability to states that would have resulted from a negotiation between two self interested agents. Thus culture affects both the states being reached and the design of the transition function. The reward function which defines the payoff an agent can expect from performing an action a in state s, could be strongly affected by culture. Agents from different cultures are expected to have different goals and hence different rewards for the same actions. For example, in Middle Eastern culture there is a high value for respect and hence high rewards associated with actions that show respect even though the action may not be beneficial monetarily.

5 Related work

There is a vast body of research for identifying the relevant psychological factors and building a theory of negotiation [21, 28]. We will briefly discuss here the literature that consider cultural effects in negotiation (for a more in-depth discussion see [10]). The effect of the cultural background of the negotiators on the negotiation process and the negotiation outcome has been studied both theoretically and experimentally in the behavioral sciences literature (see [10, 25, 6]). Cultural values and norms affect the importance people ascribe to different issues and their interpretation of the opponents behavior. Brett [5] identifies and discusses the effect of three cultural values in cross-cultural negotiation: (a) individualism versus collectivism (b) egalitarianism versus hierarchy, and (c) high versus low context communication. There is a cultural stereotype between the East and the West based on these values. A typical Western individual is presumed

to be individualistic, egalitarian and uses low context communication whereas a typical Eastern individual is collectivist, hierarchical, and uses high-context communication. However, this is a very gross characterization and cultural differences within neighboring regions also affect negotiation ([16] discusses this in the context of East Asian cultures). Although the knowledge of the opponents culture may be helpful in negotiation, there has been studies showing that negotiations can break down when negotiators adjust to their opponents culture and try to overcompensate [1]. Another important cultural factor in negotiation is cultural *sacred values* of the negotiators. People have high emotional attachment to the sacred values [2] and any act during the negotiation process perceived to violate them may result in a breakdown of the process.

There has also been some effort into building computational models of negotiation and building software agents for negotiation (see [4, 15] and references therein). However, there is relatively little work on including the effect of culture in the computational models of negotiation (except [11]). The computational models for negotiation use a variety of techniques from game theory [18, 7], probabilistic decision theory [27], bayesian learning [29], and other heuristic approaches [9]. It has been noted that the way people act in a negotiation (or in general strategic interaction) scenario do not correspond to the game-theoretic equilibria [7, 13]. The decision theory-based approaches encode the agents preferences in a utility function and choose the decision with highest expected utility. Moreover, most game-theoretic and decision-theoretic models are mainly interested in the outcome of the negotiation-game instead of the process of the negotiation. We are crucially interested in both outcome and process in modeling cultural effects in negotiation. That being said, game-theoretic techniques may still be useful in analyzing the different outcomes in negotiation due to cultural differences and this is an open research problem.

To model the effect of culture in the negotiation process, we need our model to be expressive enough to model the interactive process between the agents. The agents should also be able to maintain and update knowledge about their opponents. Therefore, in this paper, we use a POMDP for modeling the negotiation process. POMDPs have been used before in modeling human social interaction where knowledge of the opponents need to maintained [22]. More recently, POMDPs have been used in a game-theoretic setting for modeling a finite repeated game between two agents [23].

6 Conclusions and Future Work

In this paper, we first showed the mapping of a general cognitive schema for negotiation to a POMDP model. We then described the POMDP setup for a general negotiation problem and discussed the challenges in modeling a negotiation problem using POMDPs. We presented a simple example of a single-issue transactional negotiation to illustrate the POMDP formulation. While many competing techniques to model a negotiation problem exist, the POMDP based modeling has the following advantages: (a) POMDPs provide a decentralized way of solving the problem which is an inherent characteristic of the negotiation domain. (b) POMDPs provide a natural way to capture

the sequential nature of the bargaining process i.e. they capture the process rather than just focusing on the outcome. (c) POMDPs can express the various important factors that affect the negotiation such as culture. We also discussed how cultural factors can be accounted for and how they affect the POMDP modeling.

Acknowledgments: We would like to thank Nazli Mukadder Turan, Laurie Weingart, Tinglong Dai, Lingzhi Luo and Mike Lewis for their valuable inputs and discussion. This research was funded by a MURI award through ARO grant number W911NF-08-1-0301.

References

1. W. L. Adair, M. S. Taylor, and C. H. Tinsley. Starting out on the right foot: negotiation schemas when cultures collide. *Negotiation and Conflict Management Research (to appear)*.
2. S. Atran and R. Axelrod. Reframing sacred values. *Negotiation Journal*, 24(3):221–246, July 2008.
3. D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 2007.
4. P. Braun, J. Brzostowski, G. Kersten, J. B. Kim, R. Kowalczyk, S. Strecker, and R. Vahidov. *Intelligent Decision-making Support Systems*, chapter e-Negotiation Systems and Software Agents: Methods, Models, and Applications, pages 271–300. Springer London, 2007.
5. J. M. Brett. Culture and negotiation. *International Journal of Psychology*, 35(2):97–104, 2000.
6. J. M. Brett and T. Okamura. Inter- and intra-cultural negotiations: US and Japanese negotiators. *Academy of Management Journal*, 41:495–510, 1998.
7. C. Camerer. *Behavioral game theory: Experiments in stratgeic interactions*. Princeton University Press, Princeton, NJ, 2003.
8. D. Druckman. Stages, turning points and crises: Negotiating military base rights, spain and the united states. *The Journal of Confict Resolution*, 30(2):327–360, 1986.
9. P. Faratin, C. Sierra, and N. R. Jennings. Negotiation decision functions for autonomous agents. *Int. Journal of Robotics and Autonomous Systems*, 24(3-4):159–182, 1998.
10. M. J. Gelfand and J. M. Brett, editors. *The handbook of negotiation and culture: Theoretical advances and cultural perspectives*. Stanford Business Books, Stanford, CA, 2004.
11. G. J. Hofstede, C. M. Jonker, and D. Verwaart. An agent model for the influence of culture on bargaining. In *HUCOM 2008, First International Working Conference on Human Factors and Computational Models in Negotiation*, pages 39–46, 2008.
12. L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.
13. J. Kagel and A. Roth, editors. *The handbook of experimental economics*. Princeton University Press, Princeton, NJ, 1995.
14. S. Kraus. Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, 92(1-2):79–97, July 1997.
15. S. Kraus. *Strategic negotiation in multiagent environments*. MIT Press, Cambridge, MA, USA, 2001.
16. K. Leung and D. Tjosvold, editors. *Conflict management in the Asia Pacific*. Wiley, Singapore, 1998.
17. R.B. Myerson. Two-person bargaining problems and comparable utility. *Econometrica*, 45(7):1631–1637, 1977.
18. S. Napel. *Bilateral bargaining: theory and applications*. Springer, 2002.

19. P. Paruchuri, M. Tambe, F. Ordonez, and S. Kraus. Security in multiagent systems by policy randomization. In *AAMAS*, pages 273–280, 2006.
20. J. Pineau, G. Gordon, and S. Thrun. Point-based value iteration: An anytime algorithm for pomdps. In *IJCAI*, 2003.
21. D. G. Pruitt. *Negotiation behavior*. Academic Press, New York, 1981.
22. D. V. Pynadath and S. C. Marsella. Psychsim: Modeling theory of mind with decision-theoretic agents. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1181–1186, 2005.
23. D. Ray, B. King-Casas, P. R. Montague, and P. Dayan. Bayesian model of behaviour in economic games. In *NIPS*, 2008.
24. R. Simmons and S. Koenig. Probabilistic navigation in partially observable environments. In *IJCAI*, 1995.
25. C. H. Tinsley. Models of conflict resolution in Japanese, German, and American cultures. *Journal of Applied Psychology*, 83:316–323, 1998.
26. C.H. Tinsley. How negotiators get to yes: Predicting the constellation of conflict management strategies used across cultures. *Journal of Applied Psychology*, 86(4):583–593, 2001.
27. J. von Neumann and O. Morgenstern. *The Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 2nd. ed., 1947 edition, 1944.
28. L. R. Weingart, J. M. Brett, M. Olekalns, and P. L. Smith. Conflicting social motives in negotiating groups. *Journal of Personality and Social Psychology*, 93(6):994–1010, 2007.
29. D. Zeng and K. Sycara. Bayesian learning in negotiation. *International Journal of Human-Computer Studies*, 48(1):125–141, January 1998.

Culture and Negotiation: Three Models

Brosh M. Teucher[1]
Jeanne M. Brett[1]
Brian C. Gunia[1]

[1]Kellogg School of Management
Northwestern University
Evanston, IL 60208, USA
{b-teucher, jmbrett, b-gunia}@kellogg.northwestern.edu

Abstract. We propose a refinement to Brett´s model of culture and negotiation. Prior theorizing and research on the role of culture in shaping negotiation dynamics had emphasized joint gains as the central outcome of a negotiation. Brett´s [1, 2] cultural model focused on integrative strategies leading to critical insights, which can yield joint gains. Her research identified Questions and Answers (Q&A) within Western cultures and Offers within Eastern cultures as integrative strategies that maximize joint gains. These strategies were expected to be both normative and dominant in Western and Eastern cultures, respectively. Recent research by Gunia et al. [3], focusing on Indian negotiators revealed a normative distributive strategy, which dominates Q&A and Offers, and systematically produces low joint gains. The domination of this normative, distributive, and suboptimal strategy is not explained by current models of culture and negotiation. To address this anomaly, we refine Brett´s model to include the cultural dimensions of trust and mindset, which explain the process by which culture affects the deployment of negotiation strategy. We develop a 2x2 cultural classification, integrating mindset (linear or holistic) and trust (a priori trust or distrust) as cultural determinants of normative negotiation strategy. We identify four prototypes of negotiation cultures: "Arrow" (linear & trust), "Bliss" (holistic & trust), "Clash" (linear & distrust) and "Diplomat" (holistic & distrust). We propose an inter-cultural negotiation research framework capitalizing on the classification, and we discuss theoretical implications, future research and applications.

Keywords: Culture, Negotiation Strategy, Joint Gains, Trust, Mindset

Dispute Resolution Research Center (DRRC) working paper #388

1 Introduction

Joint gains maximization has been the primary criterion for evaluating negotiation outcomes since Luce & Raiffa's [4] *Games and Decisions* [see also 5, 6].There are many reasons why economists and psychologists who study negotiations emphasize joint gains maximization. Joint gains imply 1) that no deal value has been left, unclaimed on the table; 2) that tradeoffs satisfying both parties' interests have occurred; 3) agreements are

possible when without these gains there is no zone of possible agreement 4) stable agreements and harmonious, long-term relationships.
In this paper, we first explain Brett's [1, 7] model of culture's impact on intercultural negotiation, focusing especially on the negotiation strategy portion of that model, and reviewing the relevant empirical research. We then propose an elaborated model of culture, negotiation strategy, and joint gains.

1.1 Joint Gains in Intercultural Negotiations

Brett defines culture as encompassing two elements: psychological and socio-structural. The psychological elements include values, beliefs, and norms shared by members of a group, whereas the socio-structural elements include the economic, social, political and religious institutions that provide the context for intra-group social interaction. Figure 1 presents Brett's model of culture's impact on negotiation. The dependent variable in Brett's model is the negotiation outcome ("Outcome"). The model represents two negotiators: one from culture "A" and the other from culture "B". Each negotiator has a set of interests and priorities that underlie that negotiator's positions on the issues to be negotiated. These interests and priorities may be influenced by culture. For example, in Tinsley's negotiation exercise "Cobalt vs. Silverlight," [8] the Cobalt company's fundamental interests, underlying their positions about the nature of their joint venture, concern control over the joint venture. In contrast, the Silverlight Company's interests concern status and face: how the Korean government and Silverlight's industrial partners view the joint venture.

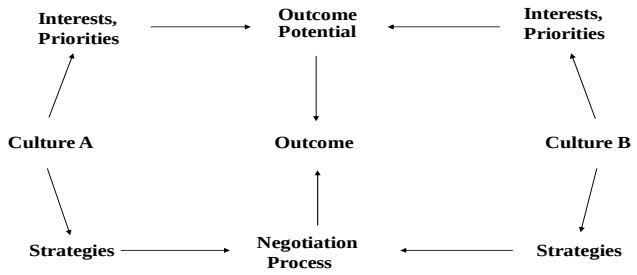

Fig. 1. Intercultural negotiations model [1]

Negotiation theory has long noted that the tradeoffs inherent in different interests and priorities generate the potential for integrative outcomes, i.e., outcomes that generate joint gains [9]. Brett's model, however, indicates that whether negotiators reach that

potential depends on the pattern of social interaction (i.e., negotiation process) that results from the strategies each negotiator brings to the negotiating table. A negotiation strategy is a set of goal-driven behaviors applied throughout a negotiation According to Brett's model, culture influences the strategies that negotiators tend to deploy. Whether intercultural negotiators will achieve joint gains, depends on the efficacy of their strategies for identifying joint gain opportunities, as well as the fit between the negotiators' strategies [2].

1.2 Negotiation Strategy and Joint Gains

The earliest psychological research on negotiation strategy and joint gains was conducted by Deutsch [10] and Pruitt [9]. Other important papers include Weingart et al. [11] and Olekalns and Smith [12]. This line of research – all of which was done in the U.S. or Australia – identifies the constructs in Figure 2 as important strategic precursors to joint gains. Figure 2 requires that we understand four more constructs: aspirations, distributive strategy, integrative strategy, and insight.

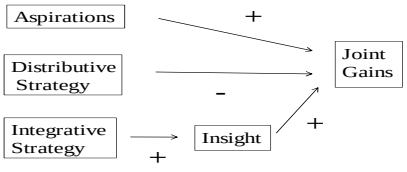

Fig. 2. Model of negotiation strategy and joint gains

Aspirations refer to negotiation goals – not just *what* but *how much* the negotiator wants to gain from the negotiation [13]. The research clearly indicates that negotiators who have higher aspirations negotiate higher joint and individual gains. Aspirations keep negotiators motivated throughout the negotiation to identify an agreement that fulfills those aspirations; they deter satisficing[1] and motivate search [9, 14]. One cultural group that consistently demonstrates high aspirations, for example, is Israelis: Brett [7] characterized the Israeli negotiators she studied as pragmatic individualists who set high targets and were motivated to search for information and willing to make trade-offs in order to accomplish their goals.

[1] For a discussion on satisficing see the work of Simon [26, 27]

Negotiation strategy is a set of goal-driven behaviors applied in a negotiation situation. Two broad classes of negotiation strategies were identified. The first, distributive strategy is behavior oriented around the goal of maximizing the individual negotiator's portion of the negotiation outcome. Research on distributive strategy [9, 15] has identified two major sets of strategic behaviors that negotiators use to claim individual gains: Substantiation (i.e., persuasion) and Offers. Commonly these behaviors are applied jointly. For short, we refer to these behaviors, which underlie distributive strategy, as S&O. A long line of research has documented that distributive strategy has a negative impact on joint gains [9, 16]. The reason seems to be that distributive behavior places the focus of the negotiator solely on self, rather than on the self and other. Moreover, distributive negotiators do not engage in the search for information, which is the core of integrative negotiation strategy.

Integrative strategy is behavior oriented around the goal of maximizing joint *and* individual gains in negotiation. The mechanism linking integrative behavior to joint gains is insight – understanding that an opportunity exists to benefit both negotiators [9, 17]. Insight occurs when negotiation strategy leads one negotiator to realize the interests and priorities of the other party. Such a realization enables the negotiator to identify possible tradeoffs between the two parties and to propose offers that capitalize on these tradeoffs.

A long line of research on Western-culture negotiators has identified Question and Answer (Q&A) behavior as components of integrative strategy that produce insight [9, 11, 12]. Negotiators can bundle knowledge generated through Q&A and the sharing of (honest) information about priorities into multi-issue offers that reflect tradeoffs. This line of research treats single issue offers as indicators of distributive strategy and multi-issue offers as indicators of integrative strategy [12], because only multi-issue offers allowed for the explicit bundling of tradeoffs.

1.3 Culture, Negotiation Strategy, and Joint Gains

We begin with the assumption based on prior research and theorizing that cultures provide commonly accepted-strategies of action expected to be effective in particular situations, like negotiations. In an early study of Japanese and U.S. negotiators' strategies and outcomes, Adair, Okumura, and Brett expected to find differences on both dimensions [2, 18]. They did *not* find differences in outcomes - joint gains of intra-cultural Japanese and intra-cultural U.S. negotiators were comparable. However, they identified very different intracultural negotiation processes. Japanese negotiators made both single - and (less often) multi-issue offers from the earliest moments of the negotiation, and continued to make offers throughout the negotiation. In contrast, U.S. negotiators initially engaged in Q&A and made few offers; only later in the negotiation did they start making offers. Thus, the study identified two different negotiation strategies, which emerged in different cultural contexts but created equivalent levels of joint gains.

Research led by Adair has generated three important conclusions that help explain this difference in negotiation processes: First, negotiators from high context communication cultures like Japan, who begin using and reciprocating offers early in the

negotiation, actually use offers to generate joint gains [19]. The high context communication style [20] used by these negotiators reflects their holistic mindset. This mindset facilitates a particular search process that can transform information contained in offers into the insights necessary for obtaining joint gains. Brett [1] explains this dynamic simply: negotiators from high context communication cultures with a holistic mindset use offers as a search mechanism.

A holistic mindset deters negotiators from finalizing an agreement based on single-issue offers, and enables them to infer the other parties' priorities from the pattern of offers and counter-offers. Thus, negotiators with a holistic mindset do not need to engage in direct Q&A to infer interests and tradeoffs. Brett points out that negotiators coming from holistic, high context cultures can be reasonably comfortable exchanging offers because of the implicit assumption that negotiators seldom outright lie when making offers. Negotiators may exaggerate, but rarely make offers at-odds with their own interests (because they run the risk of having this inferior offer actually accepted by the other side) [19].

The second conclusion from the Adair/Brett/Okumura line of research is that Q&A is the lowest common dominator for inter-cultural negotiations. Moreover, Q&A is not effective in generating joint gains in inter-cultural negotiations when one negotiator is from a high context and another from a low context communication culture [20]. The authors illustrate this in two of their studies. Adair and Brett [19] studied negotiations between Japanese-U.S. and Hong Kong Chinese-U.S. parties. Inter-cultural negotiations resulted in Q&A frequency levels indistinguishable from those obtained between two low-context negotiating parties (Figure 3). However, inter-cultural joint gains were significantly lower than those obtained by low-context negotiating parties. The reason seemed to be that the inter-cultural negotiators did not convert information obtained through Q&A into offers during the second half of the negotiation as often as the same-culture negotiators did (Figure 4).

The third conclusion is that low context negotiators have difficulty reaching insights from patterns of offers in the same way that high context negotiators do, and that high context culture negotiators have difficulty reaching insights from Q&A in the same way that low context negotiators do [19]. Another study, contrasting Japanese and U.S. negotiators, shows this best [21]: The earlier in the negotiation that U.S. negotiators made offers, the lower their subsequent joint gains were, and vice-versa for Japanese negotiators. In addition, the more Japanese negotiators engaged in Q&A before making the first offer, the lower their subsequent joint gains were – and vice-versa for U.S. negotiators (Figures 5 and 6). In sum, counter-cultural negotiation behavior was antithetical to joint gains.

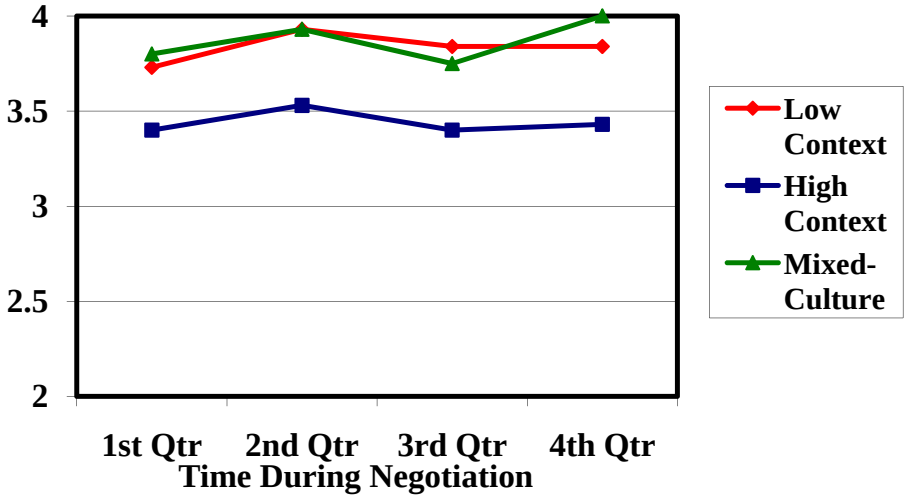

Fig. 3. Frequency of reciprocal Q & A over time [1]

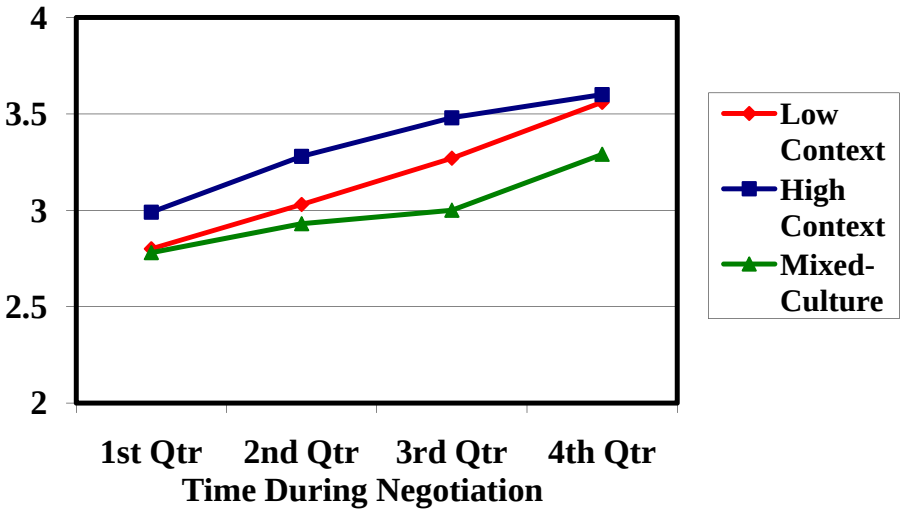

Fig. 4. Frequency of reciprocal offers over time [1]

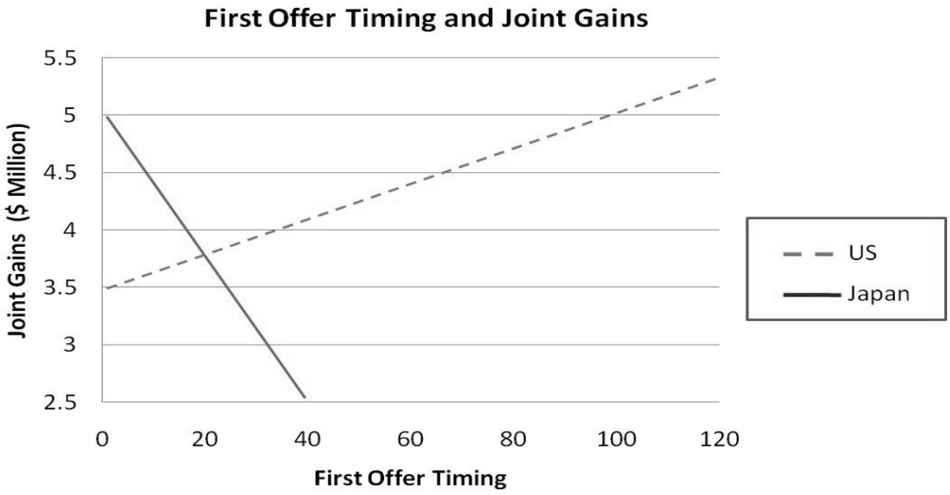

Fig. 5. U.S. and Japanese dyads' first offer timing and joint gains in intra-cultural negotiations [21]

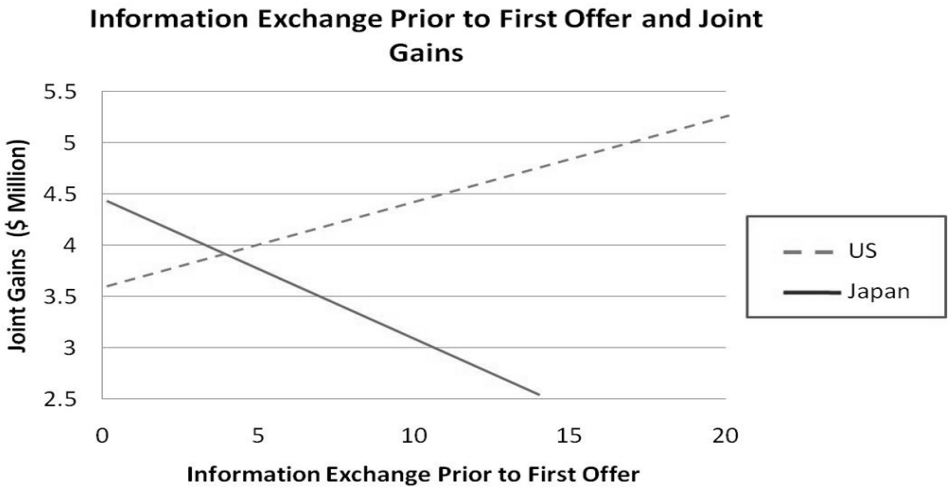

Fig. 6. U.S. and Japanese dyads' frequency of information exchange prior to first offer and joint gains in intra-cultural negotiations [21]

2 Culture, Negotiation Strategy, and Joint Gains: A Refined Model

Not all negotiators use Q & A or offers to negotiate joint gains. Negotiators generally leave gains, and negotiators from some cultures, like India, tend to leave significantly more gains than negotiators from other cultures, like Israel or the U.S. [1] Gunia, Brett, and Kamdar [3] set out to identify why. Their research showed that Indian negotiators reach lower joint gains than U.S. negotiators do, and that Indian negotiators make many more offers and substantiation attempts (S&O) and much less information sharing through Q&A than U.S. negotiators do [3]. This pattern of negotiation behaviors indicates a distributive strategy, shown previously to undermine joint gains. Mediation analyses indicated that Indian negotiators' low joint gains levels derived from this strategic pattern.

Cultural commentators are split on whether India is a high or a low context communication culture [1, 22, 23]. Following theory on culture, negotiation strategy and joint gains, if India were a low context communication culture, then Indian negotiators should use the Q&A strategy to generate high joint gains, like U.S. negotiators do. However, if India were a high context communication culture, then Indian negotiators should use the offer strategy to generate high joint gains, as Japanese negotiators do. However, the evidence is strong that Indian negotiators use S&O but generate low joint gains. The question that arises from these results is why Indian negotiators avoid the integrative Q&A strategy of U.S. negotiators, and also fail to generate joint gains through the use of offers while negotiators from other Eastern cultures, who also engage in substantiation as well as offers do so successfully [18]. This question, resulting from Gunia and colleagues' studies, motivated an examination of two matters at the heart of culture and negotiation theory: The first is specifying which element of high and low context culture influences the deployment of negotiation strategy. The second is the impact of other cultural dimensions on the deployment of strategy.

High versus low context communication varies along a number of dimensions, including directness, explicitness, logic, and ultimately (and most importantly) holistic versus linear thinking [24]. Although many elements of high versus low context communication are likely to reveal their importance as negotiation research progresses, it seems quite clear at this juncture that the holistic thinking element s central to the choice of negotiation strategy. It is this, holistic mindset that underlies Japanese negotiators' ability to turn S&O into insights capable of generating joint gains. In low context communication cultures like the U.S., where linear thinking prevails, frequent and early offers tend to generate low joint gains. The reason seems to be that linear thinking prevents negotiators from reaching broad insights from a disparate series of offers and results in resolving one issue at a time.

An equally-important question is which element of low context communication cultures allows Q&A to prevail (at least in the first half of the negotiation), produce insight captured in multi-issue offers in the second half, and ultimately generate joint gains? The answer seems to be a priori or swift trust [3]. Trust, or the willingness to make oneself vulnerable to another, in expectation of the other's positive intentions [25], seems

to facilitate Q&A. In negotiation, trust manifests itself in negotiators' exchange of information about preferences and priorities. This information could make them vulnerable to one another because each could later use it to claim value. By the same token, asking for information is pointless unless negotiators trust one another enough to believe the answers they receive. In addition, asking questions could make negotiators vulnerable by demonstrating the gaps in their knowledge.

Negotiators are understandably reluctant to make themselves vulnerable by sharing information or asking questions. However, clear cultural differences are evident, in willingness to trust [28], especially in negotiations [3]. U.S. negotiators tend to embrace swift trust in the context of negotiations, trusting until the other party proves to be untrustworthy. Not so for Indian negotiators: they appear to embrace the opposite assumption, distrusting until the other party proves trustworthy. U.S. negotiators report that they are willing to trust in negotiation, and those who do also report using more Q&A than do Indian negotiators. Indian negotiators, on the other hand, tend to embrace distrust, and those who do report significantly less Q&A and more S&O. than do U.S. negotiators. U.S. and Indian negotiators' self reports are reflected in their actual negotiation behaviors, and the pattern of their negotiation behaviors: Americans' heavier use of Q&A and Indians' heavier use of S&O predict their differing levels of joint gains [3].

These most recent studies contrasting Indian and U.S. negotiators have led us to refine Brett's theory of culture, negotiation strategy and joint gains. We suggest that culture affects constructs that are immediately important to negotiation, like trust, and mode of thinking[2]. We suggest that the way to view the interaction of culture and strategy in Brett's model, represented as Figure 1 in this paper, is to view each negotiator's strategy in terms of at least a two by two matrix (see Figure 7).

[2] We are pretty sure that status is another construct strongly affected by culture and immediately important to negotiation, at least in some cultures. However, integrating status into the model goes beyond the scope of this paper.

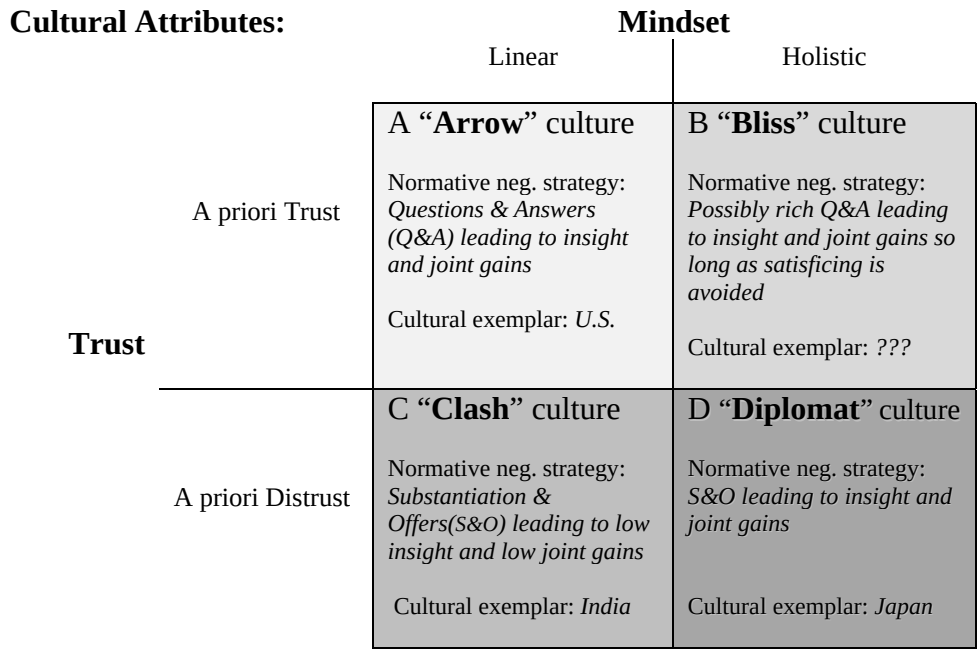

Fig. 7. A cultural model of normative negotiation strategy formation

The matrix in Figure 7 demonstrates how the cultural dimensions of trust (a priori trust, a priori distrust) and mode of thinking (linear, holistic) interact to create four normative negotiation strategies. Specifically, cell "A" or the "Arrow" culture represents a culture characterized by a priori trust and linear thinking. A cultural exemplar known to follow this pattern is the U.S. We predict that this cultural configuration will prescribe Q&A negotiation strategy, which will lead to insight and high joint gains. We call this the "Arrow" culture to reflect the explicit and well-directed Q&A exchange. Cell "B," titled "Bliss," represents a culture characterized by a priori trust and holistic thinking. We were not able to identify a cultural exemplar for this pattern. Perhaps indigenous cultures (such as Native American) that are characterized by a holistic view of nature and have had some contact with Western traditions might exhibit such patterns. If such culture does exist, we predict that it would prescribe a rich Q&A negotiation strategy that would lead to insight and high joint gains, but only if the negotiating parties did not satisfice and stop searching for the best agreement once a preliminary agreement was identified, and some of their needs satisfied. We title this culture as "Bliss" to emphasize the expected harmonic and positive negotiation interaction. Cell "C," titled "Clash," represents a culture characterized by a priori distrust and linear thinking. A cultural exemplar known to follow this pattern is India. We predict that this cultural configuration will prescribe an S&O negotiation

strategy that will lead to low insight and low joint gains. We title this cell "Clash" to signify the clash of positions that the negotiators are involved in during the S&O interaction. Cell "D," titled "Diplomat," represents a culture characterized by a priori distrust and holistic thinking. A cultural exemplar known to follow this pattern is Japan. We predict that this cultural configuration will prescribe an S&O negotiation strategy that will lead to insight and high joint gains. We call this cell "Diplomat" to emphasize the indirect and subtle insight generation process.

3 Discussion

In this paper we propose a refinement to Brett´s model of culture and negotiation. Prior theorizing and research on the role of culture in shaping negotiation dynamics emphasized joint gains as the central outcome of a negotiation. Brett´s [1] cultural model focused on integrative strategies leading to insight – the critical element for realizing joint gains. Research led by Adair identified a Question and Answer (Q&A) strategy within Western cultures and an offer strategy within Eastern cultures as normative, integrative strategies that maximize joint gains. This research concluded that these strategies would dominate in Western and Eastern cultures, respectively.

Recent research focusing on India, by Gunia et al. [3], revealed S&O as normative, as dominating Q&A, and as systematically producing low joint gains. This research concluded that Indian culture's characteristics of a priori distrust, coupled with a linear mindset (at least in the negotiation context), made S&O into a distributive negotiation strategy. The circumstances under which S&O is a distributive versus an integrative strategy are not explained by current models of culture and negotiation. To address this anomaly, we refined Brett´s model to include cultural dimensions of trust and mindset. We developed a 2x2 cultural classification integrating mindset (linear or holistic) and trust (a priori trust or distrust) as cultural determinants of normative negotiation strategy. We identified four prototypes of negotiation cultures: "Arrow" (linear & trust), "Bliss" (holistic & trust), "Clash" (linear & distrust) and "Diplomat" (holistic & distrust).

One open question in our model is the role of substantiation in the S&O strategy. Research shows that negotiators from high context cultures consistently couple offers with substantiation [18]. If offers are the key informative element exchanged in the negotiation, why do negotiators apply substantiation? Examining the role of substantiation in a "Diplomat" culture may suggest an answer. It is possible that "Diplomat" negotiators use substantiation to communicate subtle information about their priorities, e.g., information being embedded in what offers or what elements of offers are and are not substantiated. This information could help the negotiating parties craft tradeoffs without exploring priorities directly through Q&A.

This explanation could shed light on the central problem exemplified by the "Clash" culture: the low joint gains that S&O strategy produces compared to "Diplomat" culture. One explanation for this pattern is that "Clash" negotiators, who distrust their counterparts, ignore or discount the implicit information conveyed by offers,

substantiation, or both. Thus, they miss critical information that could lead to insight and joint gains. Another possibility is that although "Clash" negotiators apply S&O, as "Diplomat" negotiators do, their linear mindset limits their ability to generate insight from the patterns of information exchanged. Although compelling, these two explanations do not reconcile the inherent discord between having a linear mindset and the application of a seemingly S&O holistic negotiation strategy.

A third explanation combines the above two explanations with thoughts about substantiation. It is possible that a lack of trust compels "Clash" negotiators to open with a rigid position (offer) bolstered by substantiation. As linear thinkers, "Clash" negotiators would do well to follow with Q&A. However, Q&A requires trust. Since "Clash" negotiators do not trust, they revert to S&O. Each consecutive offer (and counter offer) is considered a revision of the previous position rather than an implicit cue about priorities. Thus, although "Diplomat" negotiators holistically evaluate the pattern of offers and substantiation, "Clash" negotiators apply a sequence of positional offers and substantiation attempts, without extracting priority information. In this manner, "Clash" negotiators apply S&O consistently with their linear mindset and low trust - a process that can eventually yield an agreement, but is very likely to leave value on the table. mutually-beneficial tradeoffs. Identifying the exact processes that differentiate "Diplomat" and "Clash" cultures is subject to future research.

The refinement of Brett's model leaves us enlightened (as it predicts how negotiators from a specific type of culture are likely to act normatively), but ultimately frustrated: the refined model still describes intra-cultural negotiations, but does not directly predict strategy and joint gains in inter-cultural negotiations. An expansion of the model in Figure 7 to the cross-cultural negotiation settings is proposed in Figure 8. Figure 8 depicts a 4x4 matrix that includes all of the possible intersections between the four basic cultural types discussed earlier. The cells located on a diagonal line from the upper left corner to the lower right corner represent the intra-cultural negotiations and the culturally-prescribed normative negotiation strategies. The questions that arise are how the different types of inter-cultural negotiations would unfold and what strategies would emerge as dominant. "Filling in the blanks" in this matrix is subject to future theorizing and research.

Finally, although Adair & Brett [19] suggest that linear, low context communication is the lowest common denominator of communication, elaborating Brett's normative model with trust suggests that linear, low context communication coupled with a priori distrust is the lowest common denominator of negotiation. This prediction has grave applied implications if one or more negotiating parties coming from different cultures default by choice or coincidence into the lowest common denominator of negotiation.

Culture "Type" / *Mindset:* / *Trust:*		**Arrow** *Linear* *A priori trust*	**Bliss** *Holistic* *A priori trust*	**Clash** *Linear* *A priori distrust*	**Diplomat** *Holistic* *A priori distrust*
Arrow	*Linear* *A priori trust*	**Normative Q&A**	? ?	? ?	? ?
Bliss	*Holistic* *A priori trust*		**Normative Q&A** **(Rich exchange)**	? ?	? ?
Clash	*Linear* *A priori distrust*			**Normative S&O**	? ?
Diplomat	*Holistic* *A priori distrust*				**Normative S&O**

Fig. 8. A partial inter-cultural model of normative and emergent negotiation strategy formation

Linear, low context communication prescribes Q&A to gain insight and generate joint gains. However, for linear thinkers, the success of the Q&A strategy is based on a priori trust. A priori distrust as a common denominator will prevent negotiators from leveraging Q&A to gain insight. Moreover, a-priori-distrust-holistic thinkers, skilled in gleaning insights from offers, will not have sufficient confidence to share and believe information to be able to utilize Q&A effectively, as evident in the research of Adair et al. [e.g., 18]. Thus, defaulting to the low context S&O negotiation common denominator will frustrate members of any one culture from capitalizing on their strengths, gain insight and generate joint gains, if the other party does not act so as to build trust. Instead low context communication coupled with a priori distrust seems likely to lead cross-cultural negotiators to revert to an S&O negotiation strategy, as illustrated by the Indian negotiators in Gunia et al.'s study.

The possibility of having the lowest common denominator as the precursor for an emergent S&O negotiation strategy, and that strategy's potential negative effect on joint gains in all of the inter-cultural cells presented in Figure 8, is ominous. This is indeed a gloomy forecast for a globalizing world where parties from different cultures are negotiating on a daily basis. Further theorizing and research is required to understand inter-cultural negotiations and how emergent negotiation strategies develop out of culturally-prescribed normative strategies. This knowledge should serve as a basis for identifying how to avoid the common denominator trap inter-culturally and negotiate joint gains.

References

1. Brett, J.M.: Negotiating globally: How to negotiate deals, resolve disputes, and make decisions across cultural boundaries. 2nd ed. The Jossey-Bass business & management series. San Francisco: Jossey-Bass. (2007)
2. Brett, J.M. and T. Okumura: Inter- and intracultural negotiation: US and Japanese negotiators. Academy of Management Journal, 41(5): p. 495—510 (1998)
3. Gunia, B.C., Kamdar, D., & Brett, J. M.: Paying a price: The consequences of cross-cultural assumptions about trust in negotiation, IACM Conference presentation, June 2009.
4. Luce, R.D., & Raiffa, H.: Games and decisions: introduction and critical survey, New York: Wiley (1957)
5. Raiffa, H.: The art and science of negotiation, Cambridge, MA: Harvard University Press (1982)
6. Walton, R.E. and R.B. McKersie: A behavioral theory of labor negotiations, New York: McGraw-Hill (1965)
7. Brett, J.M.: Negotiating globally. 1st ed., San Francisco, CA: Jossey-Bass (2001)
8. Dispute Resolution Research Center: Evanston, IL: Northwestern University (2008)
9. Pruitt, D.G.: Negotiation behavior, Cambridge, MA: Harvard University Press (1981)
10. Deutsch, M.: The resolution of conflict: Constructive and destructive processes, New Haven, Conn.: Yale University Press (1973)
11. Weingart, L.R., Thompson, L., Bazerman, M. H., & Carroll, J. S.:, Tactical behavior and negotiation outcomes. International Journal of Conflict Management, 1(1): p. 7--31 (1990)
12. Olekalns, M. and P.L. Smith: Testing the relationships among negotiators' motivational orientations, strategy choices, and outcomes. Journal of Experimental Social Psychology, 39: p. 101--117 (2003)
13. Pruitt, D.G., & Lewis, S. A.: Development of integrative solutions in bilateral negotiation. Journal of Personality and Social Psychology, 31: p. 621--633 (1975)
14. Galinsky, A.D., Mussweiler, T., & Medvec, V. H.: Disconnecting outcomes and evaluations: The role of negotiator focus. Journal of Personality and Social Psychology, 83(5): p. 1131--1140 (2002)
15. Weingart, L.R., Thompson, L., Bazerman, M. H., & Carroll, J. S.: Tactical behavior and negotiation outcomes. The International Journal of Conflict Management 1: p. 7--31 (1990)
16. Adair, W.L., Weingart, & Brett, J. M.: The timing and function of offers in U.S. and Japanese negotiations, Journal of Applied Psychology, 92(4): p. 1056--1068 (2007)
17. Thompson, L.: Information exchange in negotiation. Journal of Experimental Social Psychology, 27: p. 161--179 (1991)

18. Adair, W.L., Okumura. T., & Brett, J. M.: Negotiation behavior when cultures collide: The United States and Japan Journal of Applied Psychology, 86(3): p. 371--385 (2001)
19. Adair, W.L. and J.M. Brett: The negotiation dance: Time, culture, and behavioral sequences in negotiation. Organization Science, 16(1): p. 33--51 (2005)
20. Hall, E.T.: Beyond culture, 1st ed., Garden City, NU: Anchor Press. 256. (1976)
21. Adair, W.L., Weingart, L., & Brett, J. M.: The timing and function of offers in U.S. and Japanese negotiations. Journal of Applied Psychology, 94(4): p. 1056--1068 (2007)
22. Kumar, R.: Negotiating with the complex, imaginative Indian. Ivey Business Journal, March/April (2005).
23. Kumar, R. and V. Worm: Institutional dynamics and the negotiation process: Comparing India and China. International Journal of Conflict Management,15(3): p. 304--334 (2004)
24. Glenn, E.S. and C.G. Glenn: Man and mankind: Conflict and communication between cultures. Communication and information science. Norwood, NJ: Ablex Publishing Corporation. ii, p.334 (1981)
25. Rousseau, D.M., et al.: Not so different after all: A cross-discipline view of trust. Academy of Management Review, 23(3): p. 393--404 (1998)
26. H., A., Simon: Models of Man: Social and Rational. New York: John Wiley and Sons (1957)
27. Simon, H. A.: Rationality as a process and product of thought. American Economic Review, 68, 1--16 (1978)
28. Inglehart, R.: Modernization and postmodernization : cultural, economic, and political change in 43 societies. Princeton, N.J.: Princeton University Press (1997)

Toward a Unified Negotiation Framework: Leveraging Strengths in Behavioral and Computational Communities

Nazli Turan [†], Tinglong Dai [†,††], Katia Sycara [††], and Laurie Weingart [†]

[†]Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213

[††]Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract. While there has been a large body of negotiation literature in both Behavioral Science (behavioral) and Artificial Intelligence/Game Theory (computational) communities, there has not been an attempt to bridge the two communities to our best knowledge. In this paper, we compare and contrast the characteristics of behavioral and computational literature in negotiation. We discover that incorporating the strengths of two types of literature are valuable in expanding the horizon of research outlook.

Key words: Negotiation, multi-disciplinary research

1 Introduction

In the last few decades, negotiation has enjoyed generous research interest across disciplines from scholars who have employed methodologies and research agendas from their respective fields to unravel fundamental questions underlying bargaining situations. Political scientists have tried to reconstruct and process-analyze international negotiations between states based on official press releases from the countries (Druckman, 1986); economists have sought to understand the conditions under which negotiators equalize gains given ordinal utility scales (Myerson, 1977); policy-making researchers examined the role of equity for cooperation in international environmental negotiations (Lange and Vogt, 2003); computer scientists are also working on negotiation research, trying to design more intelligent automated negotiation agents (Sycara, 1990; Kraus, 1997). This consistent interest in negotiation speaks not only to the relevance of the concept in different facets of everyday human experience but also to its applicability as a field of scholarly inquiry. It is, however, intriguing that all these different research efforts in the area of negotiation have mostly focused on diverse views among them instead of taking advantage of the informative potential that lies in their convergence. Therefore, the main aim of this paper is to bridge the work of

two fields that have contributed greatly to research on negotiation: behavioral sciences and computational modeling. Our aim is to delineate both the similarities and differences in how these two fields have conducted negotiation research, their main findings and future directions. To our knowledge, there has not been any prior work to data aiming at an integrated framework of behavioral and computational negotiation literature. Another important purpose of this paper, therefore, is to identify future directions of negotiation research bridging the major strengths of computation and behavior researchers, especially to outline findings relating to the role of culture in negotiation.

In this paper, we have organized our thinking into four categories: utilities, consisting of values, motives and goals; agent-internal states, consisting of the more affective components of the internal reasoning of negotiators such as trust, beliefs and emotions; agent-internal reasoning, which is a cognitive process of encoding, searching and algorithm creation and finally externally-observable behavior and characteristics such as strategies, tactics and outcomes of negotiation. However, before dwelling into these four categories, it is useful to briefly present how the two fields understand and explore negotiation in general terms.

In the behavioral sciences, that is organizational behavior, psychology, social psychology, sociology and behavioral economics, one commonly used definition of negotiation is "a form of conflict behavior, which occurs when two or more parties try to resolve a divergence of interest by means of conversation" (Pruitt and Kim, 2004, pg 56). This general and broad definition perhaps demonstrates the fact that negotiation is one of the most common yet at the same time most complex human activities (Lewicki et al, 1997). People constantly engage in negotiations in their social and professional lives to solve issues that contain both shared and opposed interests (Ury and Fisher, 1981). This means that each negotiation situation contains potential for both competition and cooperation and almost inevitably mutual interdependence (Lewicki et al, 1997).

In the computational literature, there exists similar definitions of negotiation as in the behavior literature. (Braun et al, 2006) define negotiation as "a decentralized decision-making process used to search for and arrive at an agreement that satisfies the requirements of two or more parties in the presence of limited common knowledge and conflicting preferences." The research focus of most of the computational literature, however, is different in that it focuses on providing negotiation support systems (or e-Negotiation systems) to enable automated negotiations between intelligent, autonomous agents, or to design automated negotiation agents to negotiate with human counterparts, or to help and advise negotiators during the various phases during the negotiation process.

It should be noted that there is comparatively more knowledge on negotiation in the behavioral sciences since negotiation has been studied in these fields for a much longer time than it has been in computational sciences. It is also true that since behavioral sciences mainly uses experimental methods with human subjects, interactions between people have been easier to capture. Again, precisely due to these properties of the methods employed, knowledge discovered by behavioral sciences on negotiation has been more complex as it has been possible

to include or control for multiple factors such as individual differences, evolving structure of the negotiation, possibility of future interaction and so go. Therefore, it will also be very useful to take advantage of the findings from behavioral sciences to inform computational models on negotiation.

2 Negotiation Research: Different Perspectives

There exists different ways of classifying negotiation existing research. (Raiffa, 1982) propose that negotiation papers can be classified into four categories: (1) symmetrically descriptive, (2) symmetrically prescriptive, (3) asymmetrically prescriptive.descriptive, and (4) externally prescriptive or descriptive. The first category of research focuses on "describing the behavior of all the negotiators, without having any interest whatsoever in prescribing how they should behave". The second category of papers provide advices regarding how each negotiating agents should behave. Such advices are given symmetrically to all the parties. The third is concerned with "studying and understanding the behavior of real people in real conflict situations, so that he can better advise one party about how it should behave in order to achieve its best expected outcome". The last category differs from all the other three in that it addresses the problem of an interventor (e.g., mediator, arbitrator, and rule manipulator) during an negotiation scenario. The research helps interventor manage the negotiating process and achieve better efficiency.

Another way of classifying negotiation research is to look at whether a paper focuses on negotiation process or negotiation outcome. In negotiation literature, a *process* refers to the events and interactions that occur between parties before the outcome. A process includes all verbal and non-verbal exchanges among parties, the enactment of bargaining strategies and the external and situational events that influence the negotiation (Thompson, 1997). Process analysis in bargaining has mainly focused on either the back and forth exchanges between the negotiators (Adair and Brett, 2005) or on the broader phases of strategic activity over time (Olekans, Brett and Weingart, 2003). A more recent trend has been to examine the "interplay between moment-to-moment actions and reactions exhibited by negotiators within their broader behavioral/strategic context" (Olekans and Weingart, 2008).

Negotiation outcome, on the other hand, is the "product or endpoint of bargaining" such as an agreement, impasse or deadlock (Thompson, 1997). The most general categorization that comes from such analysis of negotiation outcomes and processes is the distinction between *competitive* and *cooperative* situations, which is also referred to *distributive* vs. *integrative* or *hard* vs. *soft* bargaining. Competitive negotiation occurs when "the goals of one party are in fundamental and direct conflict with the goals of the other party" and where "resources are fixed and limited and thus each party wants to maximize his own profit" (Lewicki et al, 1997). On the other hand, cooperative negotiation entails that "goals of the parties are not mutually exclusive" (Lewicki et al, 1997). Thus, in cooperative processes, parties can engage in "positive moments to increase the

potential gains relative to no agreement" whereas in competitive ones, they utilize moves that "threaten to increase the losses for the other party" (Walcott and Hopmann, 1974). Similarly, in distributive outcomes, a fixed resource is simply divided whereas in integrative outcomes, interests of both parties are satisfied although there may be concessions on both sides (Lewicki et al, 1997).

However, it is also true that it is limiting and challenging to try to divide a complex human interaction like negotiation into these two clear-cut and opposing categories. Thus, a third category, mixed-motive, is used to refer to bargaining situations where parties use a mixture or competitive and cooperative strategies to pursue their interests which usually are competing and compatible at the same time (Fairfield and Allred, 2007).

3 The Present Framework

As posited above, the present framework will present four categories: subjective utilities, agent-internal states, agent-internal reasoning and externally-observable behavior.

3.1 Utilities

Behavioral Literature In behavioral research, subjective utilities, mainly values, goals and motives in negotiation, are constructed by the individual but are also influenced by the social context that places constraints on these preferences. One of the first models that reflect this phenomenon is the Dual Concern Theory (Pruitt and Rubin, 1986). Although originally developed as a conflict management framework, issues addressed by this model are very much in tandem with those of negotiation and are thus used very often in this arena. The model outlines five basic conflict management strategies: competition, collaboration, compromise, avoidance and accommodation. An individual's choice of the strategy to employ in a conflict situation will be determined as a product of his concern for himself and concern for the other, which can be called self-concern and other-concern. For example, people high in self-concern and low on other-concern will tend to compete just as those high in both will collaborate or those low in self but high in other concern will tend to accommodate.

Most of the application of the self vs. other concern dichotomy in behavioral negotiation research has been within the framework of social motives, or social value orientations (Deutsch, 1949; Van Lange, 1999). Although these two terms are used almost interchangeably in the literature, the latter is used more commonly to refer to differences in social motives rooted in personality traits whereas the former is used both for situational and individual differences. The main dichotomy in this framework is the prosocial vs. egoistic social motive with the egoistic social motive usually being broken down into individualistic and competitive components. When applied to negotiation, it is expected that prosocials will have the aim to maximize outcomes for both self and others whereas individualistic negotiators will seek to maximize only own outcomes

and competitive negotiators will try to maximize relative advantage over others (Van Lange, 1999). In other words, prosocials draw positive value from others' outcomes whereas individualists draw none and competitors draw negative value from others' outcomes in negotiation. The implications of these propositions, especially for integrative negotiation, have been empirically tested through multiple studies. In a meta-analysis of 28 studies examining the role of prosocial vs. egoistic motives and resistance to yielding in negotiation, De Dreu, Weingart and Kwon (2000) found that "negotiators were less contentious, engaged in more problem-solving and achieved higher joint outcomes when they had a prosocial rather than egoistic motive" when resistance to yielding was not low. Findings of the authors extend previous work by demonstrating that prosocial negotiators engage in more problem-solving behavior and resort less to conscientious tactics, which is critical because it sheds light on some of the important behavioral mechanisms leading to integrative agreements.

What people value in negotiation is another broad theme that falls under the subjective utilities category. According to the *economic models of bargaining* that dominated the field in its nascent stages posit that the ultimate aim in negotiation is maximizing one's own outcome and the easiest and most efficient way to realize this aim is through integrative potential (Nash, 1953). However, it is now well-documented in the field that pure economic outcomes are poor indicators of not only what people value in negotiation but also of their behavioral manifestations. Research has shown that perceptions of self, relationship with the other party or the desire to maintain a positive image may be as influential as, if not more, than economic gains. Issues such as self-efficacy, self-esteem, maintaining face or maintaining social relationships with the other party may be of critical concern to the negotiators and subsequently influence processes and outcomes (Bandura, 1977; Synder and Higgins, 1988; Anderson and Shirako, 2008; McGinn and Keros, 2002). The question of what negotiators value and how it influences their perceptions of the outcome has become a fertile area of bargaining research to the extent that Curhan, Elfenbein and Xu (2006) developed and validated a framework to measure subjective value in negotiation. The 16-item Subjective Value Inventory (SVI) includes questions about the perception of the negotiator towards the incremental outcome in terms of satisfaction, loss/win, legitimacy; and feelings about the self, the process and the relationship. The authors also find that the SVI is a more accurate predictor of future negotiation decisions than economic outcomes, which demonstrates again that what people value in negotiation cannot be fully or accurately predicted by sole profit maximization models.

The last main sub-category to be discussed within subjective utilities is goals. At this point, it is important to recognize a possible profusion of terminology in the behavioral science perspective on negotiation. On one hand, goals are used almost interchangeably as motives and are treated in the same way that motives have been described in this paper. For example, there is considerable reference to prosocial or competitive goals in negotiation (De Dreu, 1997). On the other hand, goals are also used to refer to target or aspiration points. Most of the

work that approaches goals in negotiation from a goal-setting framework would fall under this latter group. A meta-analysis by Zetik and Stuhlmacher (2002) examined the influence of goal setting in negotiation and found a strong effect for difficult goals on higher outcomes on an individual basis.

Computational Literature In the computational literature, few papers discuss subjective utility. Rather most of the computational literature about negotiation provides a clear definition of the utility function. (Fogelman-Soulie, Munier and Shakun 1983) developed an MDP model for the problem of bilateral two-issue negotiation. Instead of assuming bivariate utilities, the one-stage payoff is expressed as a payoff probability distribution representing the probability that a player obtains various amounts of each of the two variables. (Kraus, Wilkenfeld and Zlotkin 1995) discussed different forms of continuous utility functions over all possible outcomes, e.g., time constant discount rates and constant cost of delay. (Zlotkin and Rosenschein 1996) presented an approach to the negotiation problem in non-cooperative domains wherein agents' preferences over different intermediate states are captured by "worth functions" by considering the probabilistic distance between intermediate states and final states. (Rangaswamy and Shell 1997) designed a computer-aided negotiation support system, one part of which is to help negotiating parties disaggregate their own preferences and priorities in order to have a better understanding of them, utilizing several utility assessment techniques. (Faratin, Sierra and Jennings 2002) used a given linearly addictive multi-attribute utility function to represent agent preferences. Each agent is assumed to have a scoring function that gives the score it assigns to a value of each decision variable in the range of its acceptable values. Then the agent assigns a weight to each decision variable to represent its relative importance. (Lin, et al. 2008) assumed that agents have bounded rationality and their choice preferences are modeled using the a utility function generated from the order of ranking of different offers. The analysis is shifted from the model of expected utility maximization to the evaluation of offers using the maximin method and the ranking of offers.

A number of papers, however, represent the trade-off between multiple issues using constraints instead of utility functions. (Balakrishnan and Eliashberg 1995) propose a single-issue negotiation process model where the utilities are simply the negotiation outcome, and agents' dynamic preferences are represented using a constraint with the left-hand side denoting agents' "resistance forces", and right-hand side "concession forces". (Luo, et al. 2003) consider fairness using a fuzzy constraint based model for bilateral, multi-issue negotiations in trading environments. The prioritized fuzzy constraints are used to represent trade-offs between the different possible values of the negotiation issues and to indicate how concessions should be made when they are necessary.

3.2 Agent-Internal States

Agent-internal states, consisting of the more affective aspects of negotiator behavior, such as trust, moods or emotions, have become a popular venue of re-

search in the recent years after a long-lasting reluctance to grant them a place at the bargaining table. This growing body of research has convincingly drawn attention to the importance of these concepts in shaping not only negotiators' own but also perceptions of others' states and consequently influencing bargaining behavior (Kumar, 1997; Barry and Oliver, 1996). The field has also outlines a distinction between the different types of affective states that may be in play during negotiation (or before and after as well). Affect refers to the "whole range of preferences, evaluations, moods and emotions" (Watson and Tellegen, 1985, as quoted in Thompson, 1998). Moods are "low-intensity, diffuse and relatively enduring affective states" and which influence "a whole range of social cognitions and behaviors considered to be primarily positive or negative" (Thompson, 1998). Finally, emotion "refers to the complex assortment of affects, beyond merely good feelings and bad that include several feelings of states" (Thompson, 1998). As can be understood from these definitions, affective states such as emotions and moods, though they certainly influence cognitions, are still understood as separate processes from them, which is a distinction that was not clearly delineated for many years in negotiation literature.

Most of the affect work in negotiation has been on the positive side and the general finding has been that positive emotions lead to a range of positive outcomes. Kramer and colleagues (1993) have found that when happy negotiators bargain with other happy negotiators, their individual outcomes are better. Carnevale and Isen (1986) replicate this finding at the joint outcome level and also find that happy negotiators perform better mainly because they engage in less contentious tactics and pressure their counterparts less towards concessions. On the other hand, there has been less research on the role of negative or neutral emotions in negotiation, owing partly to the relative difficulty of experimentally manipulating negative emotions such as anger, sadness, guilt or shame. Methods traditionally used to induce positive feelings such as making the participants watch happy videos before the negotiation do not work as effectively when applied to the negative realm. That being said, there is also considerable consensus in the field that negotiation with positive emotions lead to better outcomes than those with negative emotions (Forgas, 1998; Van Kleef, De Dreu, and Manstead, 2003). For example, angry negotiators have been found to claim more value (Anderson and Neale, 2007); to force their opponents into more concession making (Sinaceur and Tiedens, 2006) and to have outcome preferences with deleterious effects on the overall negotiation process (Loewenstein, Thompson, and Bazerman, 1989). Another point about work on negative emotion in negotiation is that most of it has been in comparative terms to negative emotions such as happy vs. sad or neutral negotiators. Therefore, the field is certainly open to expansion of knowledge how, under what conditions and through which mediators negative and neutral emotions influence bargaining situations.

In computational literature, there is a lack of modeling efforts in terms of subjective agent characteristics such as beliefs, prejudices, emotions and cultural factors. Most of the existing models, however, clearly specify information disclosure within negotiating agents, i.e., who knows what, who understands what.

(Kraus, Wilkenfeld and Zlotkin 1995) assume that each agent knows all relevant information about the other agent, particularly internal state and utility function over different outcomes. (Lin, et al. 2008) assume that there is a finite set of agent types. Each agent is aware of the set of possible types of the opponent but not the exact utility function. The agent has some probabilistic belief about the type of the other agent and such beliefs can be updated throughout the negotiation process. (Busch and Horstmann 2002) study a two-issue bargaining model with asymmetric information to study agent choice of how to structure bargaining. The problem involves a buyer and a seller who bargain over the price of two distinct goods (X and Y). The seller's valuation for each good is public knowledge; the buyer's valuation of X is common knowledge, but his valuation of Y is private information. The seller updates his belief each time after he receives an offer from the buyer.

3.3 Agent-Internal Reasoning

The third category, agent-internal reasoning encompasses the cognitions and the entailing behaviors that negotiators engage in as they search for information with the aim of sense-making during bargaining. Commonly referred to as the cognitive approach, this perspective posits that cognitive activity can be broken down and analyzed in terms of how negotiators process and recode abstract information. Some of the most popular subjects of bargaining research such as biases, heuristics or interpersonal attributions and judgments fall under the umbrella of this perspective before the evolution of which, the field had been more interested in untangling the basic behaviors, goals and strategies of negotiators without dwelling much into their underlying processes. Building on prospect theory and behavioral decision theory, the cognitive approach in negotiation has demonstrated strong effects of biases such as availability, anchoring and overconfidence. For instance, Galinsky and colleagues (2002) find that initial offers negotiators make serve as anchors that eventually become more influential assessments of outcome satisfaction than objective results are. In a similar way, building on information-processing theory, there have been findings on the implications of how the negotiation task or process is perceived by the negotiators. For example, Ross and Samuels (2003) demonstrated the influence of task perception on negotiator behavior with a very simple manipulation. They basically named two prisoner's dilemma scenarios that are similar in all other aspects in competitive vs. cooperative terms to find that that negotiators act in line with the names. All these findings show that how negotiators perceive the bargaining situation and the actors within the bargaining situation has the potential to exert process and outcome changing influence.

There exists a comprehensive body of computational literature of agent-internal reasoning. (Zeng and Sycara 1998) deveop an automated negotiation model wherein agents are capable of reasoning based on experience and improve their negotiation strategies incrementally. They utilize the Bayesian framework to update an agent's belief about its opponents. (Lin, et al. 2008) model an agent's internal reasoning in terms of generating and accepting offers. When